CHARACTERIZATION OF TRANSCRIPTIONAL ACTIVATORS IN THE FUNGAL CIRCADIAN CLOCK

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Kenneth Orval Gee
January 2014

CHARACTERIZATION OF TRANSCRIPTIONAL ACTIVATORS IN THE FUNGAL CIRCADIAN CLOCK

Kenneth Orval Gee, Ph. D.

Cornell University 2014

Circadian clocks are a common biological feature that allows organisms to anticipate changes in their environments by synchronizing their biological activity to the Earth's day and night cycle. These clocks are transcriptional/translational feedback loops, where transcriptional activators act as positive elements to promote the expression of negative elements, which in turn inhibit their own formation. While significant progress has been made in elucidating the function of circadian clock proteins, relatively little is known about their interactions on the molecular level.

To address this issue, I investigated the transcriptional activators White Collar (WC) I and II, which in the well characterized clock model *N. crassa* associate to form the White Collar Complex (WCC) and promote the expression of the negative clock element Frequency (FRQ). Also studied were the related WC homologs from *Phycomyces blakesleeanus*, *Trichoderma atroviride*, and *Cryptococcus neoformans*.

Bioinformatic analysis of these proteins was used to design protein

fragments based on variations in their domain architecture. Once generated, expression and purification conditions were optimized allowing for the generation of soluble variants, for crystallization screening and biophysical characterization using small-angle X-ray scattering.

Results from these studies indicate the presence of a PAS AB repeat in both *N. crassa* WC I and II. Purification of the PAS domain containing proteins with and without the proteins C-terminal zinc fingers resulted in apparent dimers in size-exclusion chromatography (SEC) experiments. Copurification of WC I and II proteins resulted in the formation of a tetrameric PAS mediated complex as shown by SDS-PAGE, SEC, light-scattering and small-angle X-ray scattering, suggesting the formation of a PAS mediated WCC containing four zinc fingers, which may facilitate the binding of multiple DNA sequences in the FRQ promoter region during transcription.

BIOGRAPHICAL SKETCH

Kenneth Gee was born to Lucinda and David Gee in LaPorte, Indiana on January 25, 1975. After graduation from LaPorte High School in 1993, Kenneth served in the United States Navy until 1997. Following this he attended Ball State University in Muncie, Indiana were he obtained a Bachelor's degree in biology and chemistry in 2003. He then attended Case Western Reserve University, where he focused on the generation of a human phage display library for the detection of radical intermediates associated with age-related macular degeneration until his graduation in 2005, with a Master's degree in Chemistry under the direction of Robert G. Salomon. From 2005 to 2013 Kenneth studied under the direction of Brian R. Crane at Cornell University where he focused on the generation and biophysical characterization of circadian clock proteins.

Dedicated to the memory of my sister,

Sarah Joyce Swisher

1976–2012

ACKNOWLEDGMENTS

I would first like to thank Brian Crane for his insight, mentoring and guidance through the years. His advice and suggestions were invaluable for the completion of the project and coupled with his extraordinarily kindness, made the years spent at Cornell truly remarkable. I will always be grateful for your efforts and it is a pleasure to know you, thank you so much for everything Brian.

I would also like to thank my committee members Rick Cerione and Steve Ealick for their support and guidance. I enjoyed watching and learning from your distinctly different professional styles and have found both excellent models for emulation.

I would like to thank all of the members of the Crane lab both past and present. In addition to making each day interesting, your thoughtful discussions and patient assistance countless times were invaluable to the project and my personal and scientific growth. In particular, I would like to thank Bhumit Patel. In addition to your friendship, your tireless efforts to facilitate my transition to industry, first by helping me obtain an internship, then later with a professional position cannot be underscored enough. Thank you Bhu. I also wish to thank Joanne Widom for her insight, advice and contributions to the project and my professional development. Without your efforts, completion of the project may not have been possible and the time spent most certainly wouldn't have been as productive.

I wish to thank members outside of the Crane group including Sharon Calhoun, Sharon De Roos and Pat Hine, James Almy, Tom Rutledge, David Neish, Larry Stull, Justin Luria, Laura Byrnes, Laura Tomasevich, Alex Hoepker, Marshall Hayes, Magali Moreau, Jonathan Alden, Jarrod French and Joe Renny for your friendship throughout the years.

Finally, I wish to thank my family for their support, love and encouragement, without which, none of this would have been possible.

Additional acknowledgments:

This work is based on research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the National Science Foundation and the National Institutes of Health/National Institute of General Medical Sciences under NSF award DMR-0936384, using the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award GM103485 from the National Institutes of Health, through its National Institute of General Medical Sciences.

Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco and is supported by National Institute of General Medical Sciences award P41-GM103311.

The project described was supported by award T32GM008500 from the National Institute of General Medical Sciences. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

“Never give in, never give in, never; never; never; never - in nothing, great or small, large or petty - never give in except to convictions of honor and good sense.”

-Winston S. Churchill

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
BLR-2	Blue-Light Regulator-2
C. neoformans	*Cryptococcus neoformans*
CDD	Conserved Domain Database
Ccgs	Clock controlled genes
CK	Casein kinase
cv	Column volume
Dmax	Maximum distance
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
FCC	FRH-FRQ Complex
FRH	FRQ-interacting RNA helicase
FRQ	Frequency
GFB	Gel filtration buffer
GFP	Green Fluorescent Protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His tag	Polyhistidine tag
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside

LB	Lysogeny broth
LOV	Light, Oxygen and Voltage
MADB	Mitotic Arrest Deficient B
MALS	Multi-angle light scattering
MES	2-(N-morpholino)ethanesulfonic acid
MOPS	3-(N-morpholino)propanesulfonic acid
N. crassa	*Neurospora crassa*
NCBI	National Center for Biotechnology Information
NLS	Nuclear Localization Signal
NSD	Normalized spatial discrepancy
OD	Optical density
P. blakesleeanus	*Phycomyces blakesleeanus*
PAGE	Polyacrylamide gel electrophoresis
PAS	Period-ARNT-Sim
PDB	Protein Data Bank
PEG	Polyethylene glycol
P(r)	Distance distribution function
Rg	Radius of gyration
RI	Refractive index
RT	Room temperature
rpm	Revolutions per minute
SAXS	Small-angle X-ray scattering

SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
T. atroviride	*Trichoderma atroviride*
TCEP	Tris(2-carboxyethyl)phosphine
Tris	Trisaminomethane
VVD	Vivid
Vp	Porod volume
v/v	Volume by volume
WC	White Collar
WCC	White Collar Complex

CHAPTER 1

GENERATION OF SOLUBLE FUNGAL CIRCADIAN TRANSCRIPTIONAL ACTIVATORS FOR BIOPHYSICAL CHARACTERIZATION

1.1 Introduction

1.1.1 Circadian Clocks

Synchronization of biological activity to the diurnal light cycle is one of the most prevalent features in nature. To facilitate this temporal organization, eukaryotic organisms have developed endogenous molecular transcriptional/translational feedback loops (TTFL), termed circadian clocks, which allow organisms to anticipate daily changes in their environment [1]. Common features among all clocks are inputs, oscillators and outputs. Inputs or environmental cues known as zeitgebers, which principally include food, light and to a lesser extent temperature, help to synchronize and entrain the clocks to the 24 h cycle [2, 3]. Oscillators are composed of positive and negative elements that form biological feedback loops (Figure 1.1). The positive clock elements are transcription factors that promote the expression of negative elements, which in turn inhibit their own activity. In addition to the negative clock elements, are outputs termed clock controlled genes (ccgs), which serve important and diverse roles in the organism's physiology [2].

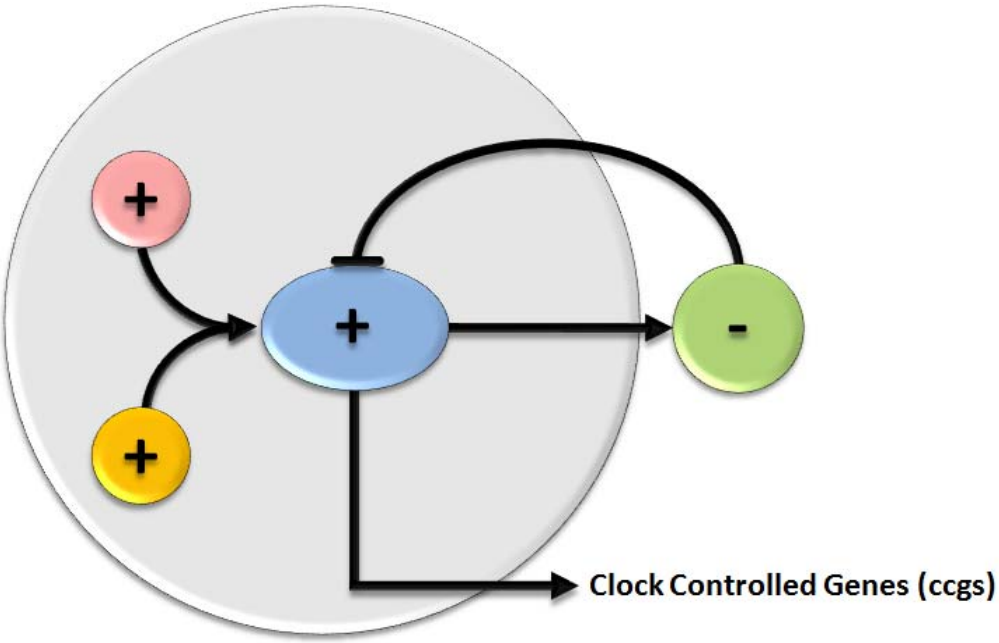

Figure 1.1 Positive and Negative Clock Element Interact to Form Biological Feedback Loops. A common feature among circadian clocks is the presence of positive and negative clock elements that form transcriptional/translational feedback (TTFL) loops. The positive elements, promote the expression of a variety of clock controlled genes including negative elements, which once expressed act to inhibit the action of the positive elements, suppressing their own transcription.

1.1.2 Using *Neurospora crassa* as a Circadian Clock Model System

To further our understanding of circadian clocks, we chose to focus on proteins derived from *Neurospora crassa* (*N. crassa*), which offers several advantages. First, *N. crassa* has a long history of study and serves as one of the best characterized clock models [4]. This understanding has come from extensive use in genetic experimentation, such as the work done by Beadle & Tatum in development of the one gene/one enzyme hypothesis, for which they were awarded the Nobel Prize [5]. Additionally, the use of race tubes (clear

glass tubes filled with growth media) has allowed researchers to readily evaluate the effects of genetic mutations on the clock by monitoring alterations in the organisms circadian mycelial (vegetative) and conidial (aerial hyphae) growth, which appear respectively as alternating dark and light bands [6]. Finally, while the individual proteins vary in the fungal, plant and animal models, the basic TTFL motif remains conserved among the different clock models. As seen in Figure 1.1.1, the positive elements in *N. crassa*, White Collar I [7] and II [8, 9], are reflected in the *Drosophila melanogaster* (*D. melanogaster*) transcription factors CLOCK and Cycle as well as CLOCK and BMAL1 in the *Mus musculus* (*M. musculus*) model [10]. Also retained is the presence of negative elements, which are Frequency in *N. crassa*, Period and Timeless in *D. melanogaster* and Period 1, 2, 3 and Cryptochrome in the *M. musculus* model [4].

1.1.2.1 *N. crassa* Circadian Clock Proteins and Their Interactions

As mentioned previously, the *N. crassa* positive clock elements are the White Collar I and II (WC I and II) PAS domain containing transcription factors. PAS domains are named for the first three proteins in which they were observed (Period, Aryl hydrocarbon receptor nuclear translocator and Singleminded) and serve a wide range of functions based on their biological context [11, 12]. The domains are composed of an α/β-fold (Aβ, Bβ, Cα, Dα, Eα, Fα, Gβ, Hβ and Iβ) with the α-helices flanking the β-sheet [11], as shown in a ribbon model of the crystal structure of *Azotobacter vinelandii* NifL (PDB: 2GJ3) (Figure 1.1.2 A) and

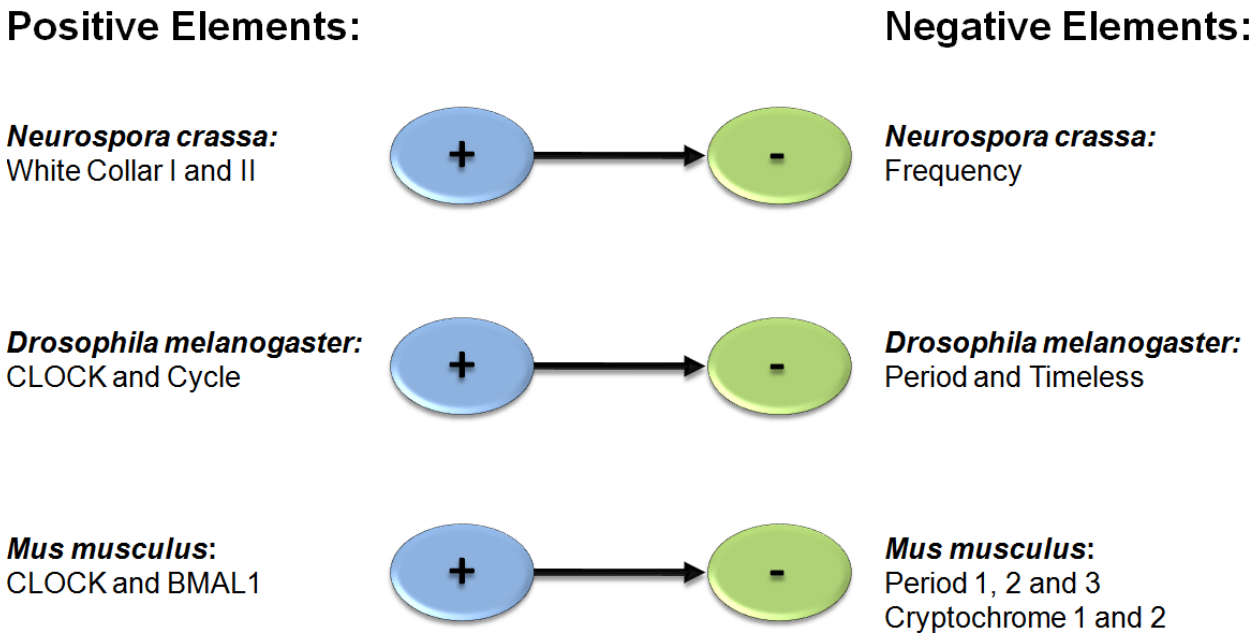

Figure 1.1.1 Positive and Negative Elements are Retained in Different Clock Models. Comparison of positive and negative elements is shown respectively from *N. crassa* (White Collar I, II and Frequency), *D.melanogaster* (CLOCK, Cycle and Period, Timeless) and the *M. musculus* model (CLOCK, BMAL1 and Period 1, 2, 3, Cryptochrome 1, 2). The presence of these proteins in different models indicates a conservation of the basic clock motif.

in a diagram of the canonical PAS fold [13] (Figure 1.1.2 B).

In *N. crassa* WC I and II, the PAS domains facilitate heterodimerization of the transcription factors starting late at night to form the White Collar Complex (WCC) [14, 15]. Additionally, they serve as the organisms photoreceptor, which helps in synchronize the clock with the diurnal cycle through the use of a flavin-binding subclass of the PAS domain, known as the Light, Oxygen and Voltage (LOV domain) in WCI [16, 17].

Following its formation the WCC binds four cis-acting pseudo-palindromic

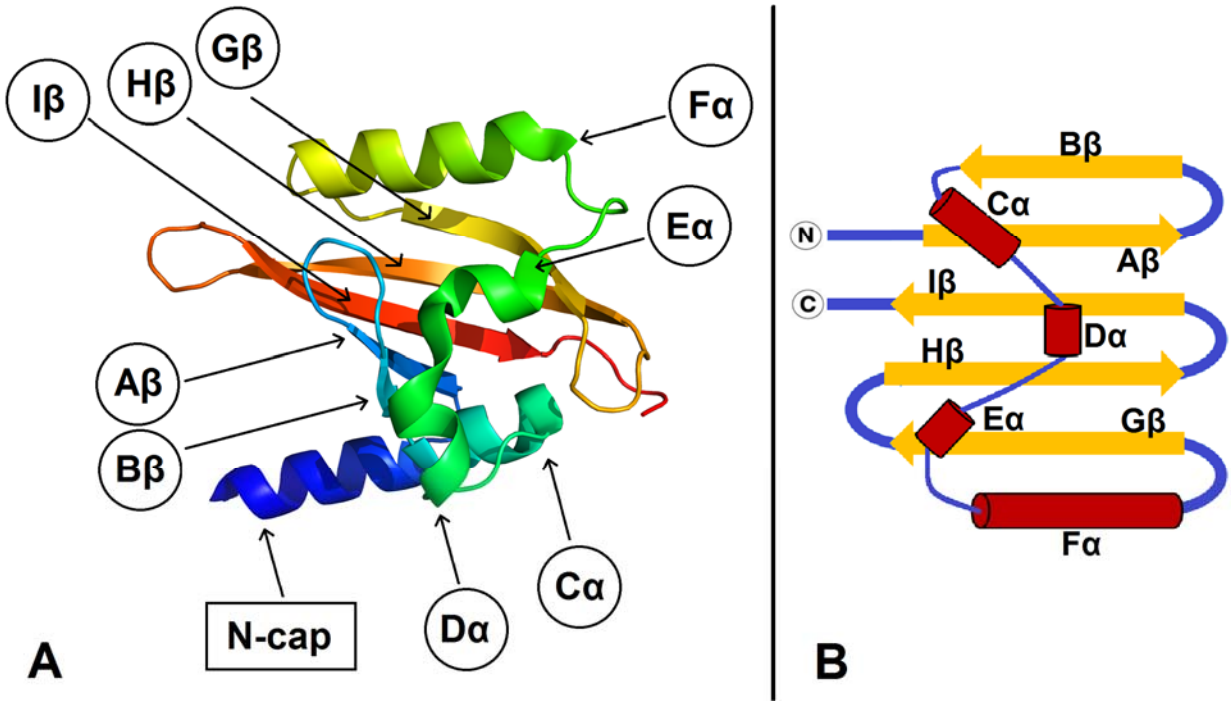

Figure 1.1.2 The Canonical PAS Domain is an α/β-fold. The PAS fold is composed of a conserved mixed α/β-fold, with α-helices flanking a β-sheet as seen in A) the ribbon model of *Azotobacter vinelandii* NifL (PDB: 2GJ3) and B) a diagram of the canonical PAS fold.

sequences termed light response elements in the *frequency* promoter region (Clock-box) [18], which upon light exposure at dawn undergoes a significant increase in activity resulting in expression of the negative element Frequency (FRQ) [18] (Figure 1.1.3). Once expressed, FRQ dimerizes though an N-terminal coiled coil domain [19] and associates with Frequency interacting RNA helicase (FRH) [20] to form the FFC [21]. The FFC then enters the nucleus where it promotes phophorylation of the WCC by casein kinase 1a (CKIa) and CK2, which decreases affinity of the WCC for the promoter, resulting in its inactivation [20-23]. Throughout the day, FRQ is progressively phosphorylated by CKIa and CK2

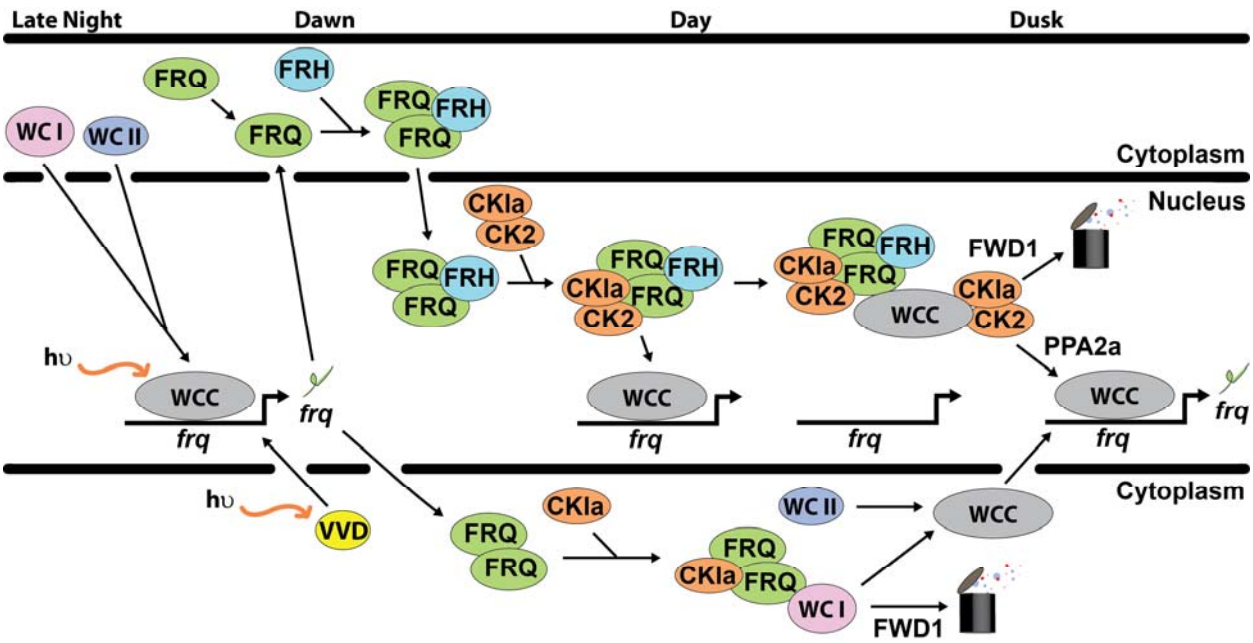

Figure 1.1.3 The *N. crassa* Circadian Clock is a Transcriptional/Translational Feedback Loop. The transcription factors White Collar I and II (WC I and II) associate to form a transcriptional activation complex (WCC), which in addition to functioning as the organisms light sensor, interacts with Vivid (VVD) to mediate light attunement. At dawn the WCC undergoes a significant increase in activity, promoting expression of Frequency (FRQ), which dimerizes and associates with a FRQ-interacting RNA helicase (FRH) to form the FFC. Once formed, FCC mediates casein kinase Ia and 2 (CKIa and CK2) phosphorylation of the WCC, resulting in its inactivation. Throughout the day FRQ is phoshorylated by CK1a and CK2, which triggers its degradation by FWD1. This decrease in FRQ mediated phoshorylation, coupled with PPA2a dephosphorylation restores WCC activity. Once bound, the active WCC is triggered for degradation by the (SCF) FWD1 Ubiquitin-ligase complex (FWD1), allowing the cycle to repeat. In addition to FRQs role at a negative clock element, is its role in stabilizing WC I in the cytoplasm prior to formation of the WCC.

[21, 24] and triggered for degradation by the Ubiquitin proteosome pathway [21, 25]. As FRQ levels decrease, the corresponding FRQ promoted WCC phophorylation also decreases and is reversed by PP2a [25]. This in turn reactivates the WCC, allowing for its binding to the promoter region and expression of FRQ, with subsequent clearance from the nucleus, as once bound

active WCC is quickly degraded allowing the cycle to restart [21, 26-28]. In addition to the core clock interactions, FRQ also participates in a positive feedback loop by mediating CKIa phophorylation of WC I in the cytoplasm following its expression (Figure 1.1.3). This serves to stabilize and inactivate the protein until it associates with WC II forming the WCC and moves into the nucleus where it is dephosphorylated by PPA2a [23].

Also observed is the interaction between Vivid (VVD) [29] and the WCC. In this case, VVD which contains a LOV domain, is thought to facilitate clock light attunement through interactions with the WCC [30] (Figure 1.1.3).

1.1.3 Approach Taken to Studying Circadian Clock Proteins

While significant progress has been made in elucidating the function [31, 32] and structures of several core circadian clock components such as the negative elements *N. crassa* Vivid [33], *Drosophila melanogaster* Cryptochrome [34] and Period [35] as well as the *Mus musculus* CLOCK/BMAL1 transcriptional activator complex [10], relatively little structural information is available to help inform the design of novel WC I and II variants (fragments of the full-length protein). This problem is further complicated by low sequence identity among PAS domains even though the α/β-fold is well conserved [36].

To address this problem, bioinformatic analysis of fungal circadian clock transcriptional activator variants from *N. crassa* WC I [37], WC II [8], as well as

the related *Phycomyces blakesleeanus* MADB [38, 39], *Trichoderma atroviride* BLR-2 [40] and *Cryptococcus neoformans* WC I and II [31, 41] was performed. From this analysis, variants based on the proteins domain architecture were generated using a bacterial expression system and purified for biophysical characterization.

Results from these efforts identified soluble PAS domain containing variants based on the inclusion of conserved residues C-terminal to the identified PAS domains, as well as variants containing the C-terminal zinc finger, allowing for crystallization screening and characterization using small-angle X-ray scattering as discussed in Chapter 2.

1.2 Materials and Methods

1.2.1 Generating Fungal Circadian Transcriptional Activators

Circadian clock transcriptional activators from *N. crassa*, *P. blakesleeanus*, *T. atroviride* and *C. neoformans* were expressed and purified for solution and structural studies. To facilitate these studies, soluble protein variants were generated from the evaluation of sequence homology to related proteins, secondary structural predictions and sequence comparison to solved structures using online bioinformatic web servers. In each case, analysis of the submitted sequence was performed using default parameters.

Comparison of the homologs primary sequence to related proteins and structures deposited in Protein Data Bank (PDB) [42] identified as having sequence homology, were facilitated using the National Center for Biotechnology Information (NCBI) [43] Basic Local Alignment Search Tool (BLAST) [44, 45]. Additional proteins for sequence comparison were identified with the Pfam domain database (Version 27.0) [46] and NCBI Conserved Domain Database (CDD) (Version 3.10) [47].

Secondary structural predictions were generated using Jpred (Version 3) [48] and XtalPred [49-52] web servers.

Primary sequence alignments were performed by GenomeNet using the ClustalW algorithm [53] and presented with ESPript [54].

1.2.2 Cloning Fungal Circadian Clock Proteins

N. crassa White Collar I and II (WC I and II), *P. blakesleeanus* Mitotic Arrest Deficient B (MADB), *T. atroviride* Blue-Light Regulator-2 (BLR-2) and *C. neoformans* WC I and II proteins were cloned into pet28 with an N-terminal 6xHis tag (Novagen) and overexpressed with *Escherichia coli* (*E. coli*) BL21 (DE3) cells (Invitrogen) using standard protocols.

DNA oligomers (Integrated DNA Technologies) and restriction enzymes (New England Biolabs) NdeI/XhoI were used to clone *N. crassa*, *P. blakesleeanus* and *T. atroviride*, whereas *C. neoformans* was cloned with NdeI/NotI. Verification of cloned DNA sequences was performed at the Cornell University Life Sciences Core Laboratories Center.

cDNA used in this study was generously provided by the following researchers: *N. crassa* White Collar I and II, by Jay C. Dunlap, Ph.D., Dartmouth College, Dartmouth Medical School, Hanover, NH. *P. blakesleeanus* MADB by Luis M. Corrochano, Ph.D., Departamento de Genética, Universidad de Sevilla, Seville, Spain. *T. atroviride* BLR-2 by Alfredo Herrera-Estrella, Ph.D., Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del IPN Sede Irapuato, Guanajuato, Mexico. *C. neoformans* White Collar I and II by Alexander Idnurm, Ph.D.,

University of Missouri-Kansas City, Division of Cell Biology and Biophysics, School of Biological Sciences, Kansas City, MO.

1.2.2.1 Cloning *N.* crassa WC I

N. crassa WC I, amino acid residues 705-918 (705-918) and (705-1000) constructs were cloned previously into pet28 with an N-terminal 6xHis tag (Novagen) using standard PCR protocols, and provided for overexpression in *E. coli* BL21 (DE3) cells (Invitrogen).

1.2.2.2 Cloning *N. crassa* WC II

Sense (forward) and antisense (reverse) primers used to clone *N. crassa* WC II proteins are shown with the N-terminal or C-terminal residue followed by the needed sequence:

Sense: 154: 5'-GCCGAC CATATG AGC ACC CTG ACC GAA TTC ACC-3', 174: 5'-GCCGAC CATATG CAG GAC TGG GAA CAC ATT CTG-3', 179: 5'-GCCGAC CATATG ATT CTG GAT GCC AAC GGA CGG-3'.

Antisense: 285: 5'-GCCGAC CTCGAG TCA GTA GGG TCG AGC CAT CAT GAA-3', 312: 5'-GCCGAC CTCGAG TCA CTC GGC AAT CCT GCG CTT CAG-3', 323: 5'-GCCGAC CTCGAG TCA CTC TTC CTG CTC CTC CTG CTC-3', 445: 5'-GCCGAC CTCGAG TCA AAT CGC GAT TCC CGC ATC ACC-3', 449: 5'-GCCGAC CTCGAG TCA TCG GTC TAG GGG AAT CGC GAT-3', 453: 5'-

GCCGAC CTCGAG TCA CGT CCG GGG ATC TCG GTC TAG-3', 455: 5'-GCCGAC CTCGAG TCA CTC TCC CGT CCG GGG ATC TCG-3', 500: 5'-GCCGAC CTCGAG TCA TTT CTT TGC CCA GCG AAG ACC-3'.

Combinations of these primers were used to generate the following *N. crassa* WC II proteins from the full-length protein, amino acid residues: 154-285 (154-285), (154-312), (154-323), (174-445), (174-449), (174-453), (174-455) and (179-500).

1.2.2.3 Cloning *P. blakesleeanus* MADB

Sense and antisense primers used to clone *P. blakesleeanus* MADB proteins are shown with the N-terminal or C-terminal residue followed by the needed sequence:

Sense: 1: 5'-GCCGAC CATATG ATG GAA TCA GCT CGT CAA CCT-3', 27: 5'-GCCGAC CATATG CCC CCA ACC ACA GTC CAG GCT-3'.

Antisense: 181: 3'-GCCGAC CTCGAG TCA ATT CTG CTGCTG TTG CTT CTG-5', 286: 3'-GCCGAC CTCGAG TCA GAC TGT TAC TCT CTC TAT ACT-5', 289: 3'-GCCGAC CTCGAG TCA TTC ATC GTG GAC TGT TAC TCT-5', 293: 3'-GCCGAC CTCGAG TCA AGT TTC TTC ATC TTC ATC GTG-5', 297: 3'-GCCGAC CTCGAG TCA CTG TTG CTC GTT AGT TTC TTC-5', 300: 3'-GCCGAC CTCGAG TCA GAC TCT CTT CTG TTG CTC GTT-5', 346: 3'-

GCCGAC CTCGAG TCA GCT TGT CTT TGC CCA ACG TAA-5', 354: 3'-GCCGAC CTCGAG TCA TAT GTC TGA CCC ATC TGG TCT-5'.

1.2.2.4 Cloning *T. atroviride* BLR-2

Sense and antisense primers used to clone *T. atroviride* BLR-2 proteins are shown with the N-terminal or C-terminal residue followed by the needed sequence:

Sense: 1: 5'-GCCGAC CATATG ATG TCC CAC GGG CCG CCT CCA-3', 133: 5'-GCCGAC CATATG CTG ACC GAG TTC ACC AAG CGC-3'.

Antisense: 297: 3-GCCGAC CTCGAG TCA TTC TGC CTC TTC TTC TCT CCG-5', 318: 3'-GCCGAC CTCGAG TCA GGA TGG AGT GAT GTC TGA CCG-5', 408: 3'-GCCGAC CTCGAG TCA AAT AGC AAT TCC GGC ATC GCC-5', 411: 3'-GCCGAC CTCGAG TCA CCG ATC CAT GGG AAT AGC AAT-5', 415: 3'-GCCGAC CTCGAG TCA TGA TCT ATT CTC CCG ATC CAT-5', 417: 3'-GCCGAC CTCGAG TCA CTC TCC TGA TCT ATT CTC CCG-5', 467: 3'-GCCGAC CTCGAG TCA GTT GCG CTT CTT TTC CTT CTT-5', 484: 3'-GCCGAC CTCGAG TTA GCC TGC AGC TGG AGT TGC CGT-5'.

Combinations of these primers were used to generate the following *T. atroviride* BLR-2 proteins from the full-length protein: (1-484), (133-297), (133-318), (133-408), (133-411), (133-415), (133-417) and (133-467).

Combinations of these primers were used to generate the following *P. blakesleeanus* MADB proteins from the full-length protein: (1-354), (27-181), (27-286), (27-289), (27-293), (27-297), (27-300) and (27-346).

1.2.2.5 Cloning *C. neoformans* WC I

Sense and antisense primers used to clone *C. neoformans* WC I proteins are shown with the N-terminal or C-terminal residue followed by the needed sequence:

Sense: 431: 5'-GCCGAC CAT ATG CCC CAT CAA ATT GAC TTC TCC-3', 456: 5'-GCCGAC CAT ATG TTG TCA TCT ACG ACT TCG TAT-3', 474: 5'-GCCGAC CAT ATG CAC CCA CTA CTA TCT CCC GCC-3'.

Antisense: 637: 3'-GCCGAC GC GGC CGC TCA AAC CAA GTC AAT TTG GAA GCC-5'.

Combinations of these primers were used to generate the following *C. neoformans* WC I proteins from the full-length protein: (431-637), (456-637) and (474-637).

1.2.2.6 Cloning *C. neoformans* WC II

Sense and antisense primers used to clone *C. neoformans* WC II proteins are shown with the N-terminal or C-terminal residue followed by the needed sequence.

Sense: 1: 5'-GCCGAC CATATG ATG TCC CTC CTC GCC GAG TCT-3', 26: 5'-GCCGAC CATATG AAT GGT GCC GCC GAG TTC CTG-3'.

Antisense: 250: 3'-GCCGAC GCGGCCGC TCA TGA TTT TTG ATC TTC ATT ACT-5', 276: 3'-GCCGAC GCGGCCGC TCA TCC TAC TTC TTCTTC CAG TTC-5', 282: 3'-GCCGAC GCGGCCGC TCA AGA ATC TTC AAC TGT CAT TCC-5', 289: 3'-GCCGAC GCGGCCGC TCA TGC AGA CTG ATT GTT GTT TAT-5', 301: 3'-GCCGAC GCGGCCGC TCA TGG AGT ATC TGT CGA TGA ATG-5', 306: 3'-GCCGAC GCGGCCGC TCA CTT GTC TCC ACC CGG TGG AGT-5', 311: 3'-GCCGAC GCGGCCGC TCA AGC TTT GTT CTT ATT CTT GTC-5', 315: 3'-GCCGAC GCGGCCGC TCA ATT CTT GTC AGC TTT GTT CTT -5', 317: 3'-GCCGAC GCGGCCGC TCA CGT TTT CGG ATT CTT GTC AGC-5', 320: 3'-GCCGAC GCGGCCGC TCA TTG AGT GTT CGT TTT CGG ATT-5', 323: 3'-GCCGAC GCGGCCGC TCA AGC AGA AAC TTG AGT GTT CGT-5', 327: 3'-GCCGAC GCGGCCGC TCA CTT GTC CCC CGA AGC AGA AAC-5', 329: 3'-GCCGAC GCGGCCGC TCA TCT CTT GTG CCC CGA AGC AGA-5', 331: 3'-GCCGAC GCGGCCGC TCA AGA CTT TTG TCT CTT GTG CCC-5', 337: 3'-GCCGAC GCGGCCGC TCA CGT TTT CGG GGG GCG TCC AGC-5', 340: 3'-GCCGAC GCGGCCGC TCA TGA AGC TCC TGC CGG GCC-5', 378: 3'-GCCGAC GCGGCCGC TCA TGC CCA TCT TAG TCC ACA GGC-5', 380: 3'-GCCGAC GCGGCCGC TCA TCG TTT TGC CCA TCT TAG TCC-5', 383: 3'-

GCCGAC GCGGCCGC TCA AGT TGA ATT TCG TTT TGC CCA-5', 392: 3'-GCCGAC GCGGCCGC TCA AGA TTG CTT ATC TTT TTT GGG-5'.

Combinations of these primers were used to generate the following *C. neoformans* WC II proteins from the full-length protein: (1-392), (26-250), (26-276), (26-282), (26-289), (26-301), (26-306), (26-311), (26-315), (26-317), (26-320), (26-323), (26-329), (26-340), (26-378), (26-380) and (26-383).

1.2.3 Expressing Fungal Circadian Clock Transcriptional Activators

Unless otherwise stated, expressed proteins were grown at 37 °C in Lysogeny broth [55] medium until the OD@600 nm was 0.7 to 1.0 while shaking at 180 to 200 rpm. In each case, the temperature was reduced to 17 °C and allowed to equilibrate for 45 min to 1 h. The cells were then induced with 100 to 200 μM Isopropyl β-D-1-thiogalactopyranoside (IPTG), expressed for 16 to 20 h and harvested with centrifugation at 4 °C for 10 min at 8,000 rpm. Following harvest, the proteins were flash frozen in liquid nitrogen and stored at -80 °C.

1.2.4 Purifying Expressed Proteins

To purify expressed proteins, approximately 25 g of frozen cellular material (cell pellets) were first resuspended and mixed on ice with lysis buffer containing 300 mM NaCl, 50 mM Tris buffer (TRIS) pH 7.5, 10 mM imidazole, 10% v/v glycerol, 1 to 2mM dithiothreitol (DTT) or 2 mM Tris(2-carboxyethyl)phosphine (TCEP) using an approximate ratio of 4 to 10 mL lysis buffer for each gram of

cellular pellet up to a maximum resuspension volume of 200 mL. The suspension was then lysed on ice with mixing between four alternating 2 s on/2 s off sonication cycles each lasting 2 min.

Following sonication, the proteins were centrifuged for approximately 45 min to 1 h at 22,000 rpm and the supernatant was first collected, then applied to an Ni:NTA Sepharose [56] gravity flow column with a 2 to 3 mL bed volume equilibrated with lysis buffer for immobilized metal affinity chromatography (IMAC) purification. Bound protein was washed with 25 column volumes [57] of lysis buffer, followed by 25 cv of a wash buffer containing 300 mM NaCl, 50 mM Tris pH 7.5, 20 mM imidazole, 10% v/v glycerol, 1 to 2 mM DTT or 2 mM TCEP. Bound protein was eluted from the column with elution buffer containing 300 mM NaCl, 50 mM Tris pH 7.5, 300 mM imidazole, 10% v/v glycerol, 1 to 2 mM DTT or TCEP (reductant composition was maintained in purifications) in 1.5 to 2 mL aliquots. Fractions containing protein as determined by Bradford assay [58], were pooled and loaded onto an ÄKTA fast protein liquid chromatography instrument coupled to a HiLoad 26/60 Superdex 200 or when specifically noted, a Superdex 75 prep grade gel filtration column (GE Healthcare) equilibrated with 150 mM NaCl, 50 mM Tris pH 7.5, 10% v/v glycerol, 1 to 5 mM DTT and/or 2-5 mM TCEP and fractionated in 6 mL aliquots using a 2.5 mL/min flow rate at 4 °C for X-ray crystallographic and small-angle X-ray scattering studies.

Fractional elutions were assessed with a Coomassie Brilliant Blue R-250 (Bio-Rad) stained 10% or 4-12% Bis-TRIS SDS-PAGE gel with a 1x dilution of MES running buffer (20x stock: 50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3) or 1x dilution of MOPS running buffer (20x stock: 50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7) (Invitrogen).

1.2.5 Concentrating Expressed Proteins

Fractions collected from size-exclusion chromatography (SEC) [59] purifications were pooled and stored at 4 °C for concentration with 10 kDa Amicon Ultra-15 Centrifugal Filter Units (Millipore). Prior to use, concentrator membranes were prepared by passing 15 to 17 mL of the gel filtration buffer (GFB) through the membrane to remove possible contaminates from the manufacturing process, such as residual glycerol on the membrane. Once equilibrated, the protein was concentrated with the filter units by centrifugation at 3,000 rpm with mixing every 10 to 20 min until the total volume was reduced to 15 to 17 mL. At this point, the centrifugation rate was reduced to approximately 2,000 rpm with mixing every 10 min until the desired concentration was obtained, as determined by Bradford assay.

Once concentrated, proteins were stored at 4 °C for immediate use or for overnight cleavage of the 6xHis tag, with varying amounts of Human α-Thrombin (Haematologic Technologies), as monitored by SDS-PAGE gel analysis.

Proteins not screened immediately were aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

SEC elution profiles and SDS-PAGE gels for each of the studied variants are presented with the N-terminal 6xHis tag.

1.2.6 Crystallization Screening of Expressed Proteins

Expressed proteins were screened manually and robotically (ArtRobbins: Phoenix) in the Weill Institute for Cell and Molecular Biology at Cornell University with custom conditions and commercially available sparse matrix screens: Crystal Screen 1 and 2, PEG/Ion Screen 1 and 2, SaltRx 1 and 2, MembFac, Crystal Screen Lite (Hampton Research) Wizard Classic 1, 2, 3, 4 (Emerald BioSystems), NeXtal Tubes JCSG Core Suite I, II (Qiagen) in hanging and sitting drops.

1.2.7 DNA Oligomers for *N. crassa* Crystallization Screening

Sense and antisense primers used to screen *N. crassa* WC II (179-500) are shown with the N-terminal or C-terminal residue followed by the needed sequence:

Sense: 5'- GATCCGCTCGATC-3’,

Antisense: 3’-CTAGGCGAGCTAG-5’.

DNA annealing was performed following a standard procedure as described on manufactures website (Idtdna.com).

1.2.8 Supplemental Data for Studied Clock Proteins

Experimental details for transcriptional activators and related clock proteins not discussed in Chapter 1 are provided in Appendix A.

1.3 Results

1.3.1 *N. crassa* WC Expression Screening

To date, efforts to express in *E. coli* soluble full-length *N. crassa* WC I and II proteins in quantities and purity needed for structural studies have proven unsuccessful in all conditions tested. Previous efforts to address this problem utilized an expression screen of White Collar proteins N-terminally fused to green fluorescent protein (GFP) based on PCR cloning of WC I and WC II truncation constructs with random primers. Production of fluorescent GFP acted as a folding reporter for the library of fusion proteins [60]. Results from these efforts identified soluble *N. crassa* WC II variants containing the Per-ARNT-Sim (PAS) [11, 61] and zinc finger domains [62].

Identification of these regions was facilitated with sequencing performed at Cornell University Life Sciences Core Laboratories Center (data not shown). From this work, *N.* crassa WC II (179-500) was selected for further analysis.

1.3.2 Initial Purification of *N.* crassa WC II (179-500)

N. crassa WC II (179-500) was expressed by growing *E.* coli at 37 °C in LB medium while shaking at 200 rpm until the OD@600 nm was approximately 1.1. The cells were then induced with 500 μM IPTG, expressed for 3.5 h at 37 °C, centrifuged for 10 min at 4 °C and 8,000 rpm, harvested and flash frozen in liquid nitrogen.

To purify the protein, 25 g of frozen cell pellets were first resuspended on ice with lysis buffer containing 500 mM NaCl, 25 mM HEPES buffer pH 7.5, 5 mM imidazole, using an approximate ratio of 4 mL lysis buffer for each gram of cellular pellet. The suspension was then lysed on ice with mixing between three alternating 2 s on/2 s off sonication cycles lasting 2 min each.

Following sonication the protein was centrifuged for 1 h at 4 °C and 22,000 rpm. The supernatant was then collected and applied to an Ni:NTA Sepharose gravity flow column with a 2 mL bed volume equilibrated with lysis buffer for immobilized metal affinity chromatography (IMAC) purification at room temperature (RT). Bound protein was washed with 5 cv of lysis buffer, followed by 5 cv of a wash buffer containing 500 mM NaCl, 25mM HEPES pH 7.5, 20 mM imidazole. Protein was eluted from the column with elution buffer containing 500 mM NaCl, 25mM HEPES pH 7.5 and 200 mM imidazole in 1.8 mL aliquots. Eluted fractions containing protein, as determined with the Bradford assay, were loaded onto a HiLoad 26/60 Superdex 200 prep grade gel filtration column equilibrated with 150 mM NaCl, 50 mM Tris pH 7.5 and eluted in 6 mL aliquots at 2.5 mL/min.

SDS-PAGE gel analysis of elution fractions from IMAC purification show a 38 kDa molecular weight band consistent with the predicted molecular weight for a *N. crassa* WC II (179-500) 6xHis tag labeled monomer (Figure 1.3.1).

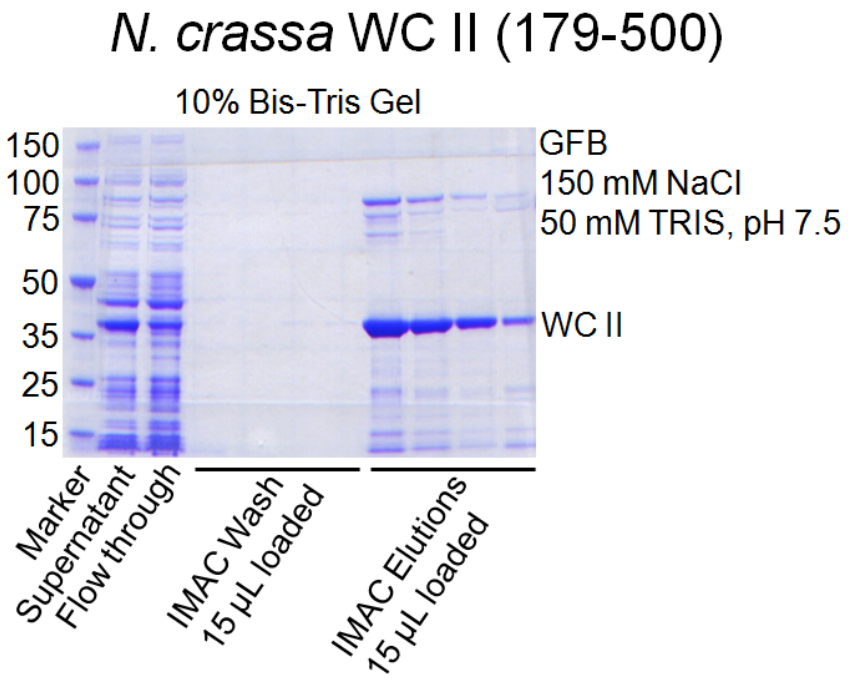

Figure 1.3.1 IMAC Purification of *N.* crassa WC II (179-500) Shows Expression of Soluble Protein. SDS-PAGE gel analysis of IMAC elution fractions shows a band consistent with the proteins theoretical 38 kDa molecular weight.

The SEC profile for *N.* crassa WC II (179-500) expressed at 37 °C indicated low absorbance peaks eluting at 192 and 220 mL, which when compared to known standards, were suggestive of a dimeric and monomeric species with apparent molecular weights of 90 and 38 kDa respectively (Figure 1.3.1.1).

This estimated molecular weight, based on elution volume, is larger than the calculated 76 kDa theoretical molecular weight for an expected 6xHis tag labeled WC II (179-500) dimer. Additionally an elution peak was also observed eluting near the void volume (120 mL) with an apparent molecular weight of

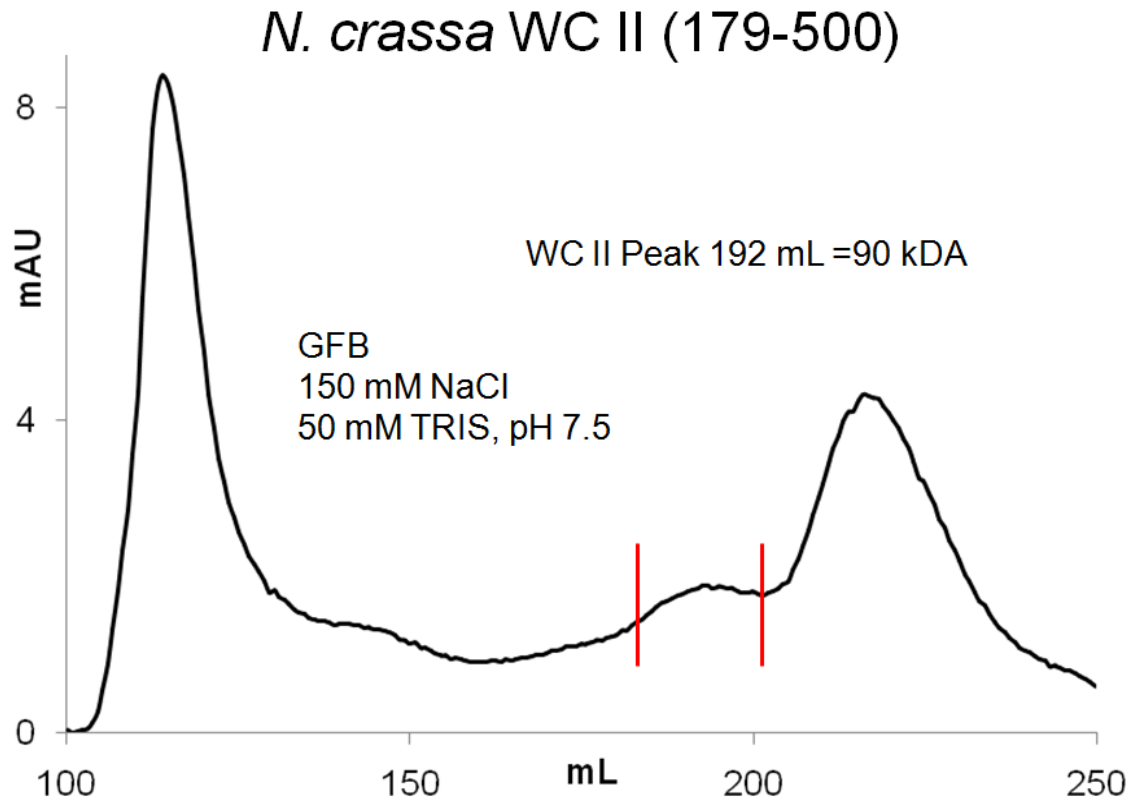

Figure 1.3.1.1 SEC Purification of *N.* crassa WC II (179-500) using Initial Conditions Shows Low Levels of Soluble Protein. The SEC profile of IMAC elution fractions shows peaks with low absorbance values at 192 and 220 mL. From comparison with elution values to known standards, these peaks indicate an apparent 90 kDa species, suggestive of a WC II (179-500) dimer (bracketed with red bars) and 38 kDa monomer. Also observed is a peak eluting near the void volume which is indicative of high molecular weight soluble aggregates.

approximately 900 kDa, suggestive of soluble aggregates.

1.3.2.1 *N. crassa* WC II (179-500) Small-scale Expression Test

In an effort to increase expression of soluble protein, a small-scale expression test was performed to determine the effect of temperature on protein solubility. To facilitate this, protein was expressed by growing *E. coli* at 37 °C in

100 mL LB medium until the OD@600 nm was approximately 1.0 while shaking at 200 rpm. The culture was then divided and one aliquot was induced with 500 μM IPTG. Cultures with and without induction were grown at RT and 37 °C with 1 mL sampling at 3 and 5 h time points. Samples were centrifuged at 4 °C for 10 min at 13,000 rpm, harvested and flash frozen in liquid nitrogen for later analysis.

To evaluate protein expression, cell pellets were chemically lysed using BugBuster Protein Extraction Reagent (Millipore), incubated for 20 min at RT with rocking and then centrifuged at 4 °C for 10 min at 13,000 rpm. 15 μL of the supernatant and 1 μL pellet resuspended in 300 μL water was then loaded on a 10% SDS-PAGE gel with MOPS running buffer.

Results from the small-scale expression testing (Figure 1.3.1.2) showed an increase in soluble protein when the induction temperature was reduced from 37 °C to RT. From this result, protein handling conditions were systematically varied in an attempt to improve protein solubility and purity.

1.3.2.2 Improved *N. crassa* WC II (179-500) Solubility

Systematic variation of *N. crassa* WC II (179-500) expression and purification conditions was performed to improve quantity and purity of expressed protein for crystallization screening. Effects on soluble protein production and stability were evaluated by observing changes in the SEC elution profile, banding pattern on SDS-PAGE gels and total protein obtained after concentration, as

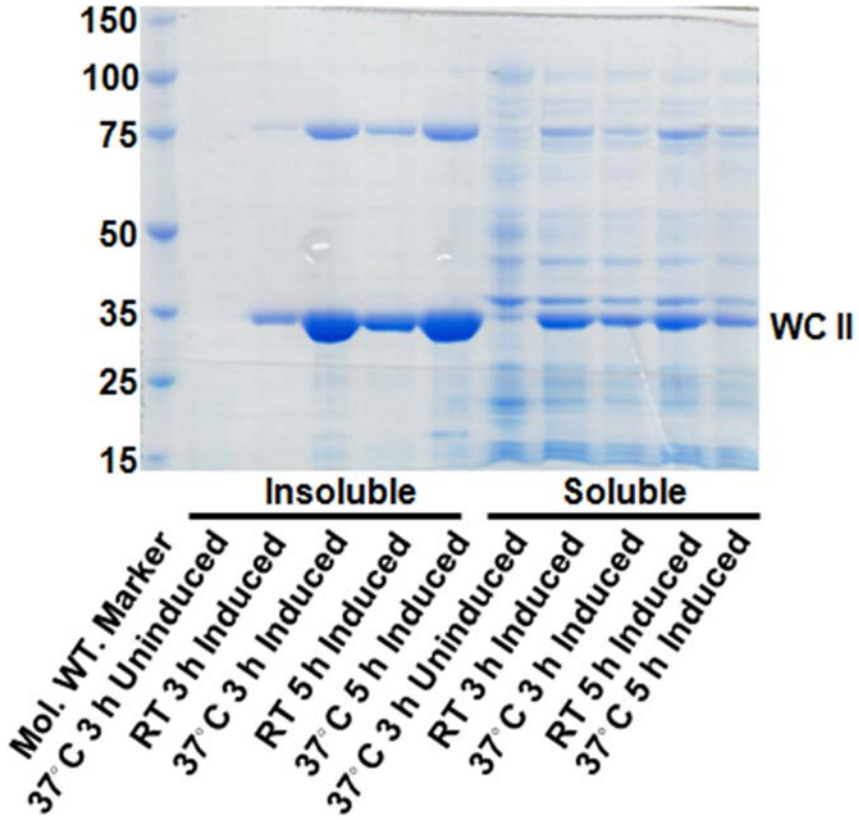

Figure 1.3.1.2 Small-scale Expression of *N.* crassa WC II (179-500) Indicates a Larger Fraction of Soluble Protein at Reduced Induction Temperatures. SDS-PAGE gel analysis of protein obtained from soluble and insoluble fraction at 3 and 5 h time points show a decrease in total protein, however, a relative increase in soluble to insoluble protein is observed when the induction temperature is reduced from 37 °C to room temperature.

quantified by the Bradford Assay.

Analysis of the SEC profile focused on absorbance intensity, shape and resolution for each of the elution peaks. SDS-PAGE gel analysis of IMAC and SEC elution fractions were evaluated for protein loss in purification steps as well as for target enrichment relative to protein fragmentation (low molecular weight bands) and irreversible aggregation (high molecular weight bands). Total protein

obtained from preparation of each cellular pellet was quantified using the Bradford assay. Conditions varied in expression steps were a reduction in the IPTG concentration from 500 μM to 100-200 μM, a decrease in induction temperature from 37 °C to RT followed by a 1 h hold time, which allowed cell cultures to equilibrate at RT prior to induction. Changes in purification steps included increasing sonication from three to four 2 min rounds (2 s on/2 sec off), increasing the resuspension volume to 5 mL lysis buffer for each gram of cell pellet and increasing imidazole concentration in the lysis buffer to 10 mM. Additionally, 10% glycerol and reductants (1-2 DTT mM or TCEP) were added as well as maintaining protein at 4 °C after thawing cellular pellets.

Results from optimization of protein handling conditions for *N. crassa* WC II (179-500) showed an increase in absorbance for the primary SEC elution peak (Figure 1.3.1.3). However as seen in the inserted SDS-PAGE gel, protein concentrated to 13 mg/mL displayed significant low and high molecular weight banding, indicative of protein fragmentation and formation of irreversible soluble aggregates even after heating for 10 min at 95 °C.

Densitometric analysis of the SDS-PAGE profile of concentrated WC II (179-500) with Syngene GeneTools indicated the 38 kDa band composed 47% of the total protein, well below the purity ideally needed for X-ray crystallization studies [63].

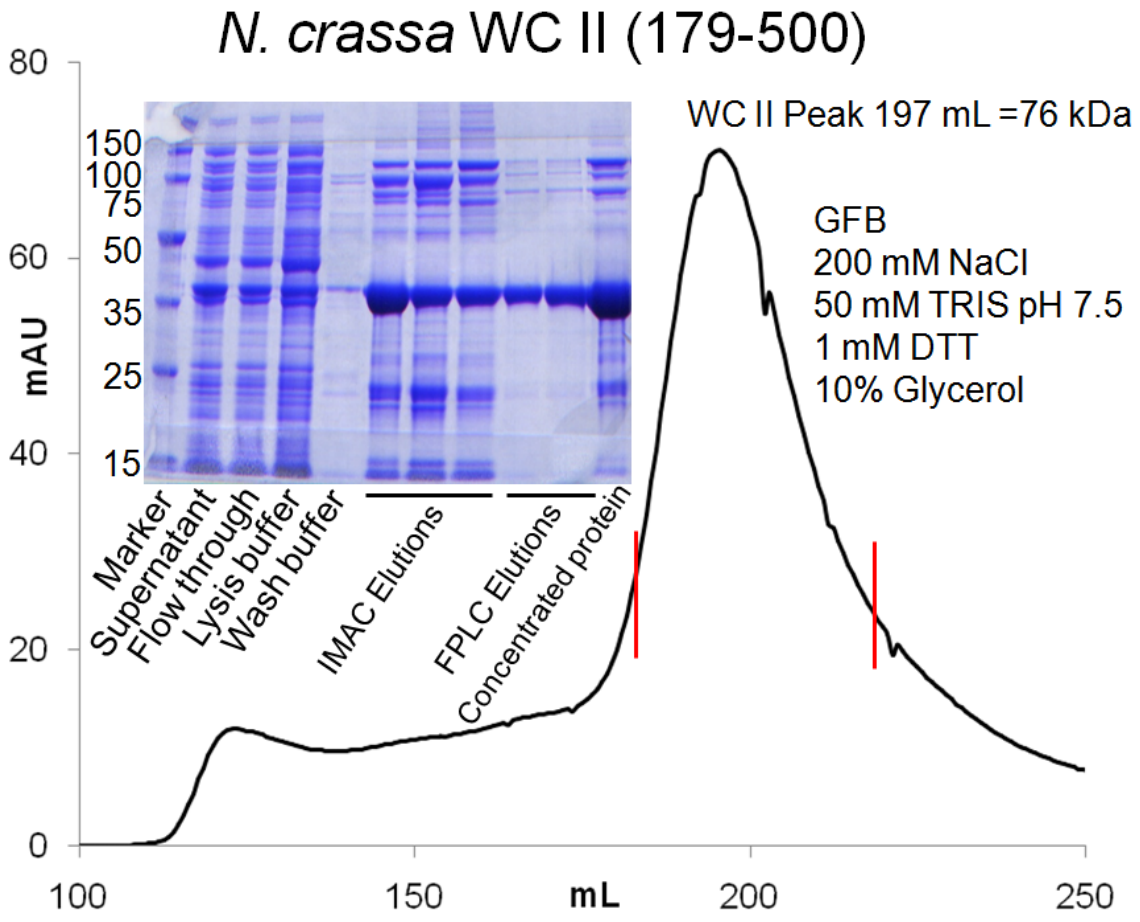

Figure 1.3.1.3 An Increase in Soluble *N.* crassa WC II (179-500) was Obtained From Variation in Expression and Purification Conditions. The SEC elution profile shows a defined strong absorbance peak eluting at 197 mL, suggestive of a species with an apparent molecular weight of 76kDa, close to the theoretical 78 kDa molecular weight expected for a dimer. SDS-PAGE analysis of SEC elutions and protein concentrated to 13 mg/mL shows protein corresponding to the 38 kDa theoretical molecular weight as well as low and high molecular weight bands consistent with fragmentation and irreversible aggregation. Red bars on the SEC elution profile denote fractions collected for concentration.

1.3.2.3 Crystallization Screening of *N. crassa* WC II (179-500)

Efforts to crystallize *N.* crassa WC II (179-500) included manual and robotic screening of the protein at a variety of concentrations using custom

conditions and commercially available sparse matrix screens. All screens were checked daily for 1 week, then weekly for 1 month and once per month (for several months).

For manual screening, the hanging drop method was used with Crystal Screen 1 and 2, PEG/Ion Screen 1 and 2, SaltRx 1 and 2, MembFac, Crystal Screen Lite (Hampton Research). Additional screening was performed with Wizard Classic 1, 2, 3, 4 (Emerald BioSystems). Drops were set up with protein ranging from 0.5 mg/mL to 14 mg/mL in varying ratios of protein:well conditions at 4, 17 °C and RT. Manual screening of custom conditions also included grid screens based on formulations found in the commercial screens with and without the addition of DTT and TCEP in concentration ranging from 1 to 5 mM in hanging and sitting drops, with 2 to 4 and 8 µL total drop volume respectively.

Screening was also conducted with rationally designed DNA oligomers based on imperfect GATA (GATN) type binding sites indentified with electrophoretic mobility shift assays [18]. Sense and antisense primers were mixed in equimolar ratios in 150 mM NaCl, 50 mM Tris buffer pH 7.5, placed in a 1 L beaker containing water, heated to 95 °C and allowed to cool gradually to RT. Once annealed, DNA was mixed with the protein in ratios ranging from 1:1 to 4:1 (DNA:protein for 30 min on ice prior to setting up crystal screens.

Robotic screening was performed with protein concentrated to 5 and 7

mg/mL at RT using Crystal Screen 1 and 2, PEG/Ion Screen 1 and 2, MembFac and Crystal Screen Lite (Hampton Research). Additional screening was performed with Wizard Classic 1, 2, 3, 4 (Emerald BioSystems) and NeXtal Tubes JCSG Core Suite I, II (Qiagen).

N. crassa WC II (179-500) concentrated to 14 mg/mL was also screened with Vivid (VVD) [64] concentrated to 18 mg/mL. Prior to screening, WC II/VVD proteins were each diluted to 7 mg/mL, mixed in a 1:1 ratio and allowed to incubate on ice for 1 h.

All robotic screens were set up using a 1:1 ratio of protein to well solution with a 1 µL total drop volume at RT. Screens were checked every 4 days for three weeks, then once monthly for several months.

Inspection of the manual and robotic screens showed that the vast majority of drop conditions (>96%) resulted in a heavy (dark gray/brown) precipitate relative to clear drops for higher concentrations screened. This ratio was observed to significantly decrease with decreasing concentration and the addition of a reductant (TCEP or DTT), resulting in a larger fraction drops containing medium grey precipitant.

1.3.3 Approach Taken to Identifying New Variants

As seen in Figure 1.3.1.3, while methods to improve the soluble expression of *N. crassa* WC II (179-500) were successful, SDS-PAGE gel

analysis showed relatively low protein enrichment due to the formation of irreversible high molecular aggregates and protein fragmentation. This observation coupled with the failure to obtain a crystal from screening efforts, underscored the need for new variants.

With this in mind, design of soluble transcriptional activator variants for *N. crassa*, *P. blakesleeanus*, *T. atroviride* and *C. neoformans* incorporated a variety of bioinformatic methods to evaluate domain-based identification, sequence conservation and secondary structural predictions for identification of potential truncations in the proteins domain architecture [65-67]. Once identified, variants were systemically PCR cloned, expressed in *E. coli* and purified for structural determination.

For evaluation of the primary sequence, a comparison of related sequences was made to those identified using the NCBI protein-protein Iterated Basic Local Alignment Search Tool (blastp), Conserved Domain Database (CDD) and the Pfam domain database. Returned values from the CDD searches are reported values from Specific Hits in the database. For each of the submitted sequences, the Pfam query returned significant Pfam-A hits, which are generated from a manually annotated library that acts as a seed during database searches [65].

Secondary structural predictions were obtained from the Jpred 3 web

server, which uses the Jnet algorithm [68] to predict α-helices, β-strands and coils from submitted sequences. Additional secondary structural predictions of α-helices and β-strands, low-complexity and disordered regions, as determined by DISOPRED2 [69] and SEG [52], were provided by the XtalPred web server. Another feature offered by XtalPred is the comparison the target proteins biochemical and biophysical features to the TargetDB [70], which is a target registration database for tracking submitted progress made worldwide on structural genomics projects. From these comparisons, XtalPred generates probability distributions and a prediction of the proteins crystallizability.

1.3.4 *N.* crassa WC II Primary Sequence Analysis

As a first step in identifying potential truncation points in the *N. crassa* WC II protein, the full-length sequence (1-530) was submitted to Pfam and CDD for analysis of the domain architecture. From this search, both databases identified a PAS fold and GATA type zinc finger domain from homology to database entries. Pfam identified residues for PAS (186-263) and zinc finger (468-503) domains from comparison to, returned Pfam closely matched CDD output, which returned (186-255) and (467-497) respectively. Also observed was a conserved Nuclear Localization Signal (NLS) from (456-461) comprising residues KKKKIK [71]. Conserved residues were identified from the alignment of non-redundant protein sequences using a blastp search on the NCBI web site. ClustalW alignment of selected sequences indicated the presence of conserved residues

N- and C-terminal to the identified PAS domain (not shown). This conservation is reflected in an ESPript web server alignment of homologs studied with *N. crassa* (numbered), where partially conserved residues are denoted with red letters in a blue box and strictly conserved residues as white letters on a red background (Figure 1.3.1.4). From these alignments, a conserved sequence N-terminal to the identified PAS domain was revealed from (154-174).

A BLAST search of structures deposited in the PDB using the Protein Data Bank proteins(PDB) database, failed to identify structures with homology to the conserved N-terminal residues. However, the search returned PAS structures with sequence identity (22-40%) to the C-terminal residues within alignments that ended between residues Ile250 to Ile355.

To further aid in the design of new variants, the primary sequence was submitted for secondary structural predictions. Results from Jpred3 analysis indicated residues likely to form α-helices (red) and β-strands (yellow) both N- and C-terminal to the identified PAS domain (Figure 1.3.1.5). Also predicted was an extended helical region (284-323) followed by alternating α-helices and β-strands extending of Ile437.

Secondary structural predictions from XtalPred were similar to results obtained from Jpred3, with an α-helix starting at the N-terminal side of the PAS domain (154-172), as well as C-terminal secondary structures for (381-389) and

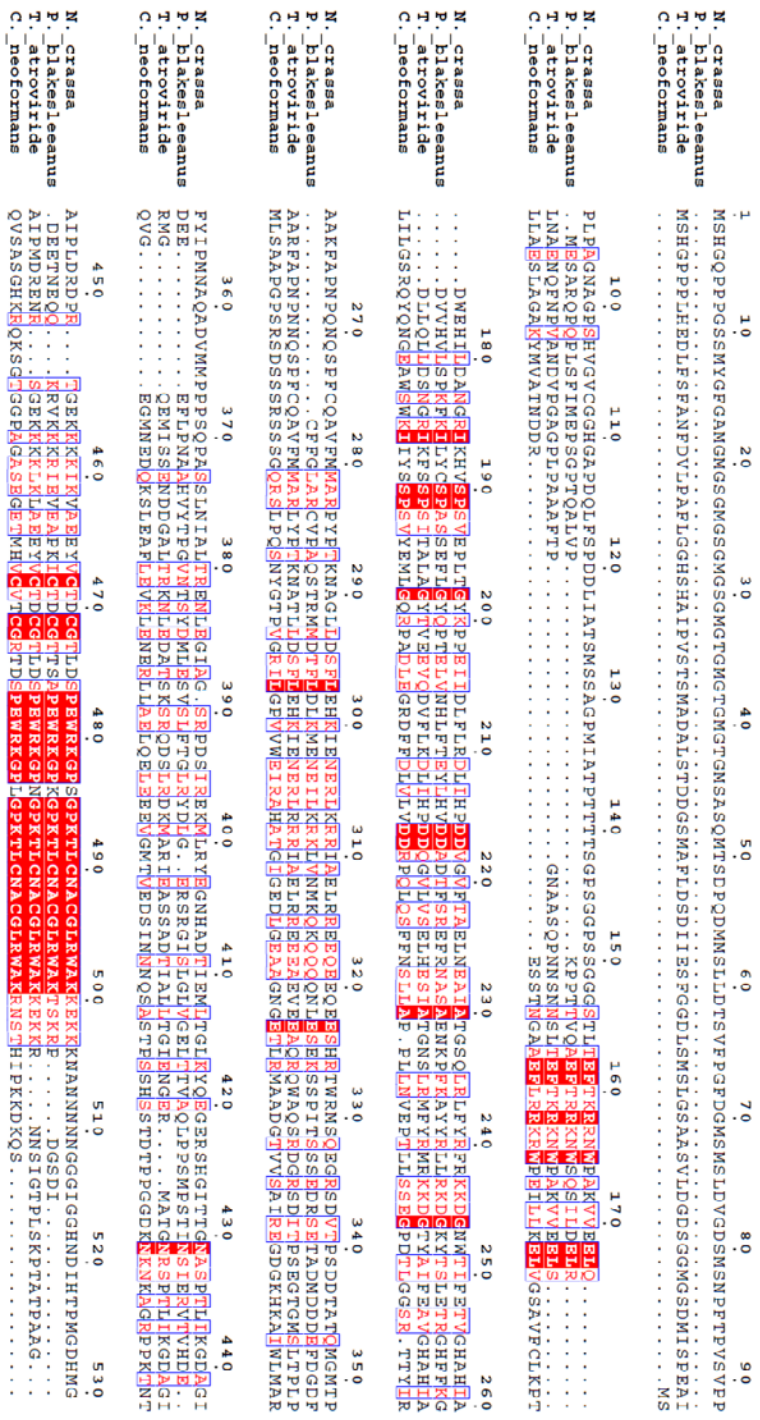

Figure 1.3.1.4 ClustalW Alignment of *N. crassa* WC II with Studied Homologs. An alignment of the full-length *N. crassa* WC II (numbered), *T. atroviride* BLR-2, *P. blakesleeanus* MADB and *C. neoformans* WC II primary sequences shows partial (red font) and strict conservation (white font with red highlighting) starting Ala96, N-terminal to the Pfam identified PAS domain (186-263) and continuing to the C-terminally end of the zinc finger (468-503).

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80...
MSHGQPPPGSSMYGFGAMGMGSGMGSGMGSGMGTGMGTGMGTGMSASQMTSDPQDMMSLLDTSVFPGFDGMSMSLDVGDSMSN

.*...90....*..100....*..110....*..120....*..130....*..140....*..150....*..160....*.
PFTPVSVPPPLPAGNAGPSHVGVCGGHGAPDQLFSPDDLIATSMSSAGPMIATPTTTTSGPSGGPSSGGGSTLTEFTKRRNWP
---EE-------------------------HHHH----HHHHHH----------------------HHHHHHHHHHHHHHHHH

.170....*..180....*..190....*..200....*..210....*..220....*..230....*..240....*..25
AKVVEELQDWEHILDANGRIKHVSPSVEPLTGYKPPEIIDLFLRDLIHPDDVGVFTAELNEAIATGSQLRLFYRFRKKDGNWT
HHHHHH---EEEEEE---EEEE--HHHHH------HHHHHHHH--------HHHHHHHHHHHHH----EEEEEEEEE----EE

0....*..260....*..270....*..280....*..290....*..300....*..310....*..320....*..330..
IFETVGHAHIAAAKFAPNPQNQSPFCQAVFMMARPYPTKNAGLLDSFLEHKIENERLKRRIAELRREEQEEQEESHRTWRMSQ
EEEEEEEEEEE-------------EEEEHHHH-HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---------

..*..340....*..350....*..360....*..370....*..380....*..390....*..400....*..410....*
EGRSDVTPSDDTATQMGMTPFYIPMNAQADVMMPPPSQPASSLNIALTRENLEGIAGSRPDSIREKMLRYEGNHADTIEMLTG
-------------------------------------------EEEE--------------HHHHHH--------HHHHHH

..420....*..430....*..440....*..450....*..460....*..470....*..480....*..490....*..5
LKYQEGERSHGITTGNASPTLIKGDAGIAIPLDRDPRTGEKKKKIKVAEEYVCTDCGTLDSPEWRKGPSGPKTLCNACGLRWA
--------------------EE---------------------------EEEE-----------------------HHHH

00....*..510....*..520....*..530
KKEKKKNANNNNNGGGIGGHNDIHTPMGDHMG
HHHHHH-------------------------
```

Figure 1.3.1.5 Jpred3 Analysis of the Full-length *N. crassa* WC II Primary Sequence. Secondary structural analysis of *N.* crassa WC II predict α-helices (red) and β-strands (blue) starting at Asp114 and continuing C-terminally to the Pfam identified zinc finger (468-503). Also observed is an extended helical region predicted immediately N- and C-terminal to the identified PAS domain (186-263).

(410-416). XtalPred also evaluated the submitted sequence for low-complexity and disordered regions using SEG [52] and DISOPRED [69] respectively. These results indicated a region of low-complexity (312-323) (italicized), a coiled coil region (296-323) (bold) and an almost continuous region of disorder from Glu317 to the C-terminal zinc finger (underlined) as seen in Figure 1.3.1.6.

To take advantage of the XtalPred crystallizability prediction feature, the primary sequence starting at residues Ser154 and Gln174 was systematically extended to include each of the conserved residues and predicted secondary

Figure 1.3.1.6 XtalPred Analysis of *N. crassa* WC II (179-500). Secondary structural analysis of the full-length primary sequence predict α-helices (red), β-strands (blue), as well as a low-complexity (itializied) (312-323) and disordered (underlined) regions both N- and C-terminal identified PAS core. Analysis also suggests the presence of a coiled coil (bold) region (296-323) C-terminal to the Pfam identified PAS domain (186-263).

elements C-terminal to the PAS domain. Crystallizability scores returned from this prediction screen (not shown) were unchanged for fragments starting at Ser154 of Gln174, however, differences were observed when the C-terminal end was varied. From these submissions, XtalPred returned the highest crystallizability prediction scores (suboptimal to average) for fragments ending between residues Met280 to Glu320 and the lowest scores (very difficult) for longer fragments up to Asp450.

From bioinformatic evaluation of the *N. crassa* primary sequence, three new variants (154-285, 154-312, 154-323) were generated for expression testing (Figure 1.3.1.7). The starting residue for the variants was based on incorporation of conserved residues predicted to form an α-helix N-terminal to the start of the canonical PAS fold (Aβ, Bβ, Cα, Dα, Eα, Fα, Gβ, Hβ and Iβ) [11, 72]. Selection of the C-terminal ending residues were based on partitioning of an extended C-terminal α-helix while minimizing the inclusion of residues predicted in low-complexity and disordered regions.

1.3.4.1 *N. crassa* WC II PAS Variants

In each case, 25 g of cellular material expressing the variant proteins was purified using conditions previously established for *N. crassa* WC II (179-500). Total soluble protein in purification steps for each of the three variants was quantified using SDS-PAGE gel analysis and the Bradford assay.

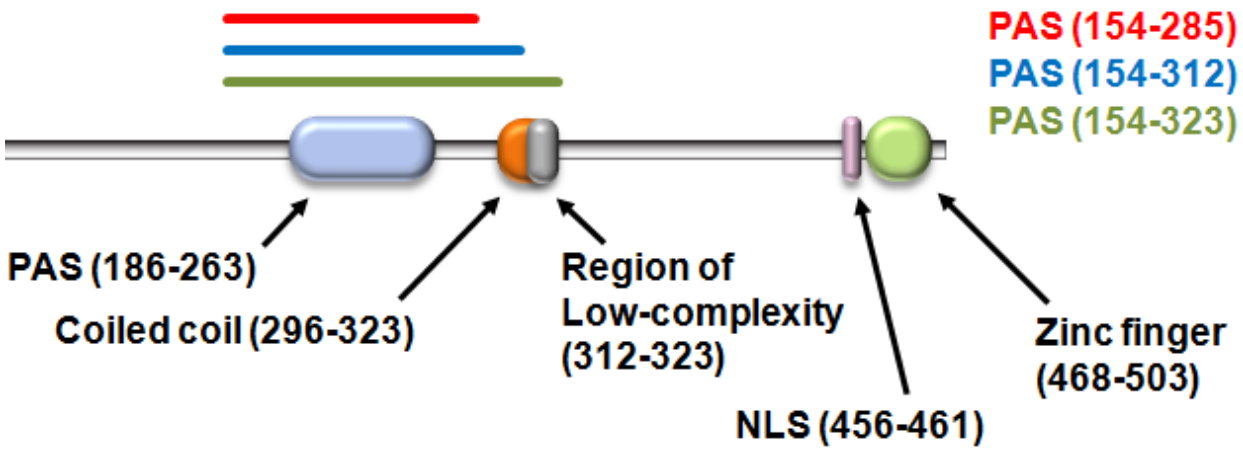

Figure 1.3.1.7 Summary of *N. crassa* WC II PAS Domain Containing Variants. Colored bars illustrate inclusion of a bioinformatically predicted PAS (186-263), coiled coil (296-313), low-complexity region (212-323), while omitting the C-terminal NLS (456-461) and zinc finger (468-503).

Results from analysis of *N.* crassa WC II (154-285) and (154-323) IMAC elutions indicated low soluble protein expression (not shown). Significantly higher levels of soluble protein were observed with (154-312), however the protein lacked stability in solution at 4 °C as it repeatedly precipitated (Figure 1.3.1.8). The first precipitation occurred during the IMAC elution and then again after 15 min at 4 °C. After each precipitation, the protein was centrifuged for 10 m at 13,000 rpm and supernatant removed. Following the second precipitation, the supernatant was subjected to SEC purification, which showed multiple low absorbance peaks and lacked a well defined peak corresponding to the expected elution value for an apparent dimer (Figure 1.3.1.8).

SEC purification of (154-312) using a Superdex 75 prep grade column also

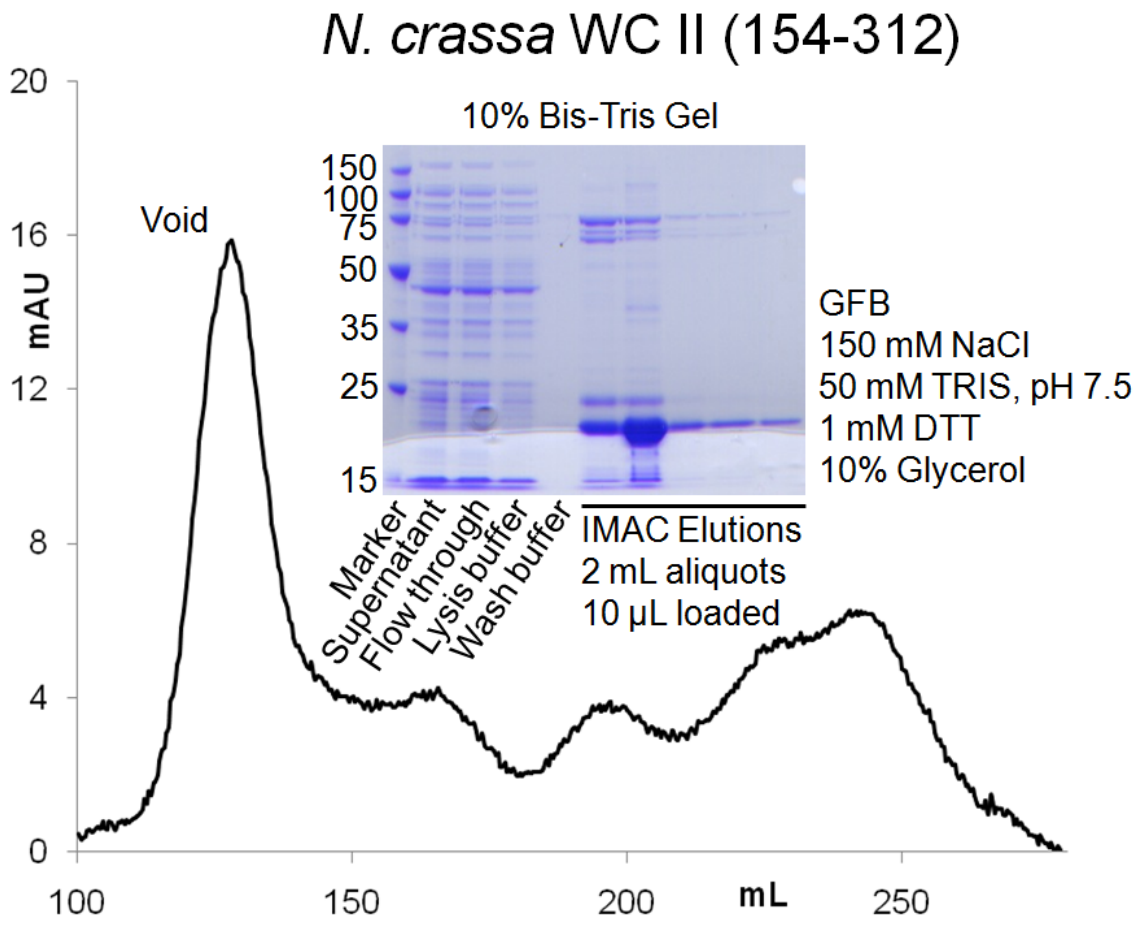

Figure 1.3.1.8 *N.* crassa WC II (154-312) Displays Low Solubility Following IMAC and SEC Purification. SDS-PAGE gel analysis of IMAC elution fractions show bands corresponding to the protein theoretical 20 kDa molecular weight, however, the protein precipitated immediately after IMAC elution. The SEC elution profile of supernatant from centrifugation of pooled IMAC elution fractions following precipitation indicates the absence of a defined elution peak consistent with the expected molecular weight for a 38 kDa dimer. A low absorption peak eluting at 243 mL is suggestive of a species with an apparent 19 kDa molecular weight, which corresponds to the value expected for a monomer.

showed multiple peaks in the elution profile with a peak corresponding to the elution value expected for an apparent monomer (not shown), similar to the elution profile, as seen in Figure 1.3.1.8.

Given the low soluble protein expression observed in variants terminating near the end of the predicted PAS domain, an alternative approach was undertaken to extend fragments from Gln174 up to the NLS. Inclusion of these residues, which might provide interactions needed to stabilize the protein, resulted in the generation of several new truncation constructs (residue ranges: 174-445, 174-449, 174-453 and 174-455) for study (Figure 1.3.1.9).

Expression and purification of 25 g preparations followed methods described for *N.* crassa WC II (179-500) with the following exceptions: addition of 5 mM DTT and TCEP, reduction of the induction temperature to 13 °C with a 1 h equilibration before induction. These changes resulted in robust soluble protein expression for each of the tested variants. SDS-PAGE analysis of SEC purification showed minimal formation of irreversible high molecular weight aggregates, however, as seen in Figure 1.3.2, low molecular weight bands attributed to fragmentation were observed.

To screen each of the generated *N. crassa* WC II PAS variants for crystallization, a similar approach as described for (179-500) was undertaken. Manual screening was conducted with Crystal Screen 1 and 2, Wizard Classic 1, 2, 3, 4 and custom grid screens using protein at stock concentration as well as those dilute to 50 and 25% using GFB at 4 and 17 °C. For each of the tested conditions, a ratio of 2 µL protein to 2 µL well condition was used. Stock protein

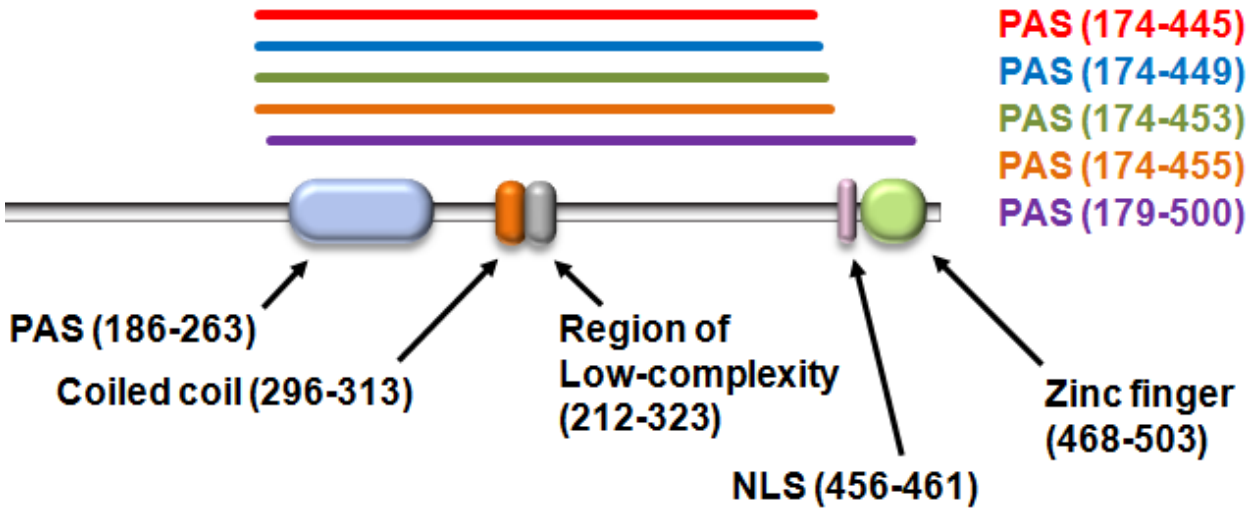

Figure 1.3.1.9 Summary of *N. crassa* WC II PAS and Zinc finger Domain Containing Variants. Colored bars illustrate inclusion of a bioinformatically predicted PAS (186-263), coiled coil (296-313), low-complexity region (212-323). Also shown is *N. crassa* WC II, which contains the predicted PAS domain, coiled coil, low-complexity region, nuclear localization signal (NLS) (456-461) and C-terminal zinc finger (468-503).

was robotically screening using Hampton Crystal Screen 1 and 2, PEG/Ion Screen 1 and 2, MembFac, Crystal Screen Lite, Wizard Classic 1, 2, 3, 4 and NeXtal Tubes JCSG Core Suite I, II using a ratio of 0.5 µL protein to 0.5 µL well condition at RT.

1.3.5 *N. crassa* WC I Variants

N. crassa WC I variants containing the C-terminal PAS (PAS C) domain (705-918) and PAS C with the C-terminal zinc finger domain (705-1000) [62] were provided by co-workers for study (Figure 1.3.2.1). WC I (705-918) was generated following expression and purification methods described for *N. crassa* WC II (174-455). SDS-PAGE gel analysis of SEC elution fractions from 25 g

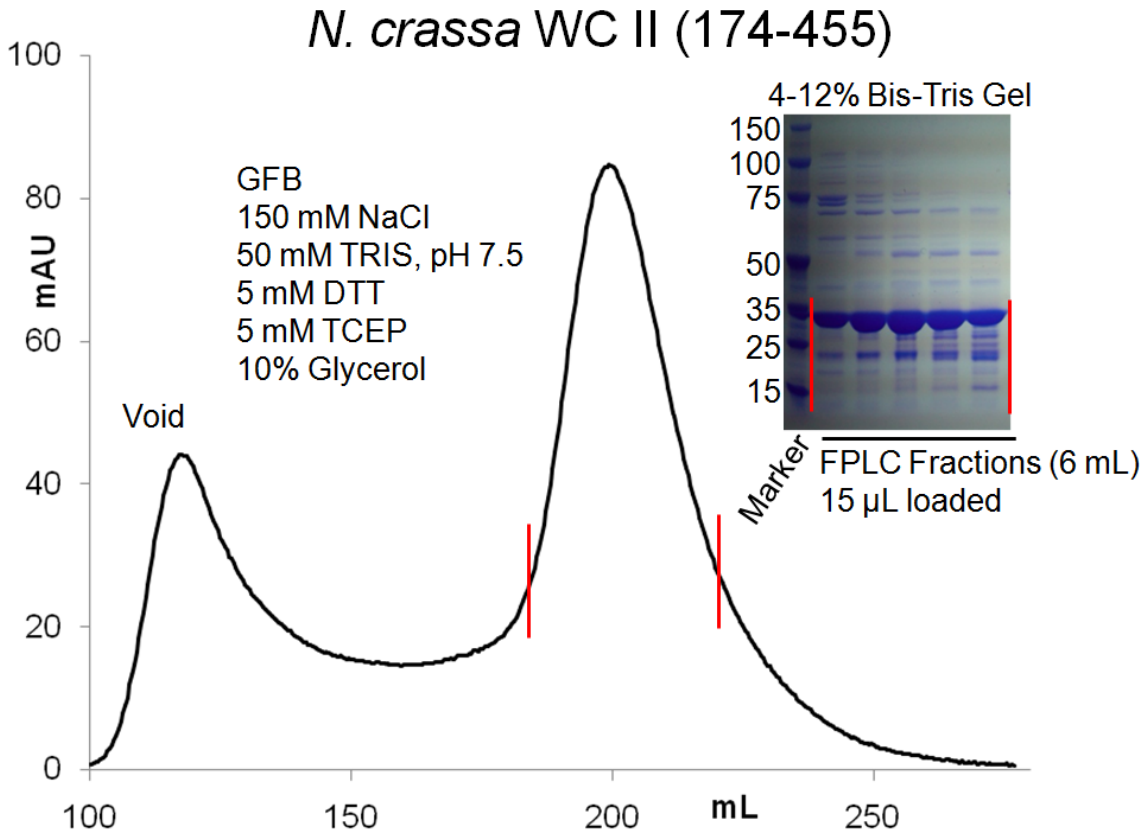

Figure 1.3.2 Extension of PAS Variants Shows Improved Stability in Solution. SDS-PAGE gel analysis of *N. crassa* WC II (174-455) SEC elution fractions show bands corresponding to the proteins theoretical 34 kDa molecular weight. A principle SEC elution peak at 199 mL is suggestive of a dimeric species with an apparent molecular weight of 72 kDa. Eluted protein selected for concentration and crystallization screening are denoted with red bars on the SEC elution profile and corresponding SDS-PAGE gel.

preparations showed significant soluble protein expression with minimal high and low molecular weight banding (Figure 1.3.2.2). Selected fractions from SEC purification were collected and concentrated with samples obtained at 2, 4 and 6 mg/mL.

N. crassa WC I (705-918) was screened for crystallization at 2, 4, and 6

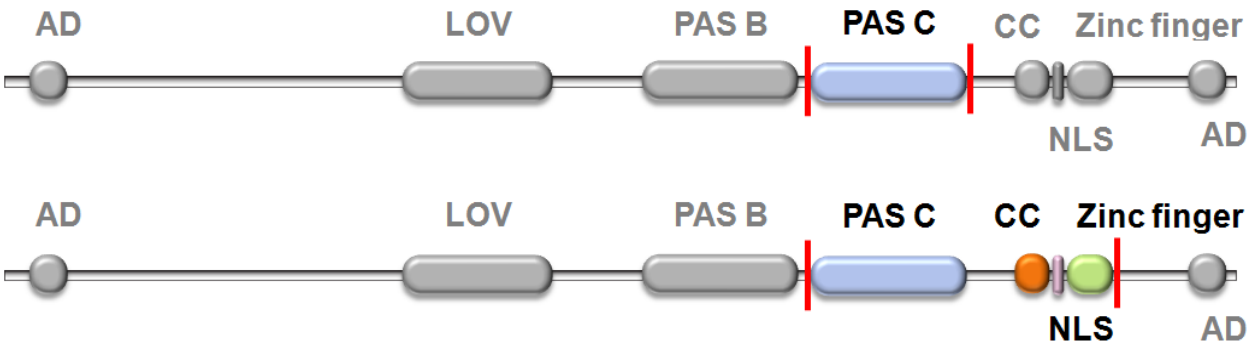

Figure 1.3.2.1 Summary of *N. crassa* WC I Domain Architecture. The WC I (705-918) variant contains the C-terminal PAS domain (PAS C) while WC I (705-1000) includes the PAS C, coiled coil domain (CC), nuclear localization domain (NLS) and zinc finger (bottom), with both variants omitting the LOV and putative activation domains (AD) bracketed with red bars.

mg/mL using Crystal Screen 1, 2 and Wizard Classic 1, 2, 3, 4 at 4 °C with 2 and 4 µL total drop volumes. Screens were checked daily for three weeks, with the majority of conditions showing a transition from medium to heavy precipitation with increasing concentration.

To screen *N. crassa* WC I (705-918) and WC II (174-455) for crystallization as a complex, stock WC I at 6 mg/mL was mixed in an equimolar ratio with WCII and tested using Crystal Screen 1 at RT. Upon addition to the well conditions, the protein immediately precipitated. The solution was then diluted with an equal volume of GFB and showed the same result. Protein was then diluted to 0.75 and 1.5 mg/mL with GFB then screened at 4 °C using Crystal Screen 1 and 2, Crystal Screen Lite and Wizard Classic 1, 2, 3, 4 in hanging drops with 2 and 4 µL total drop volumes. Additional screening was conducted with variation in the

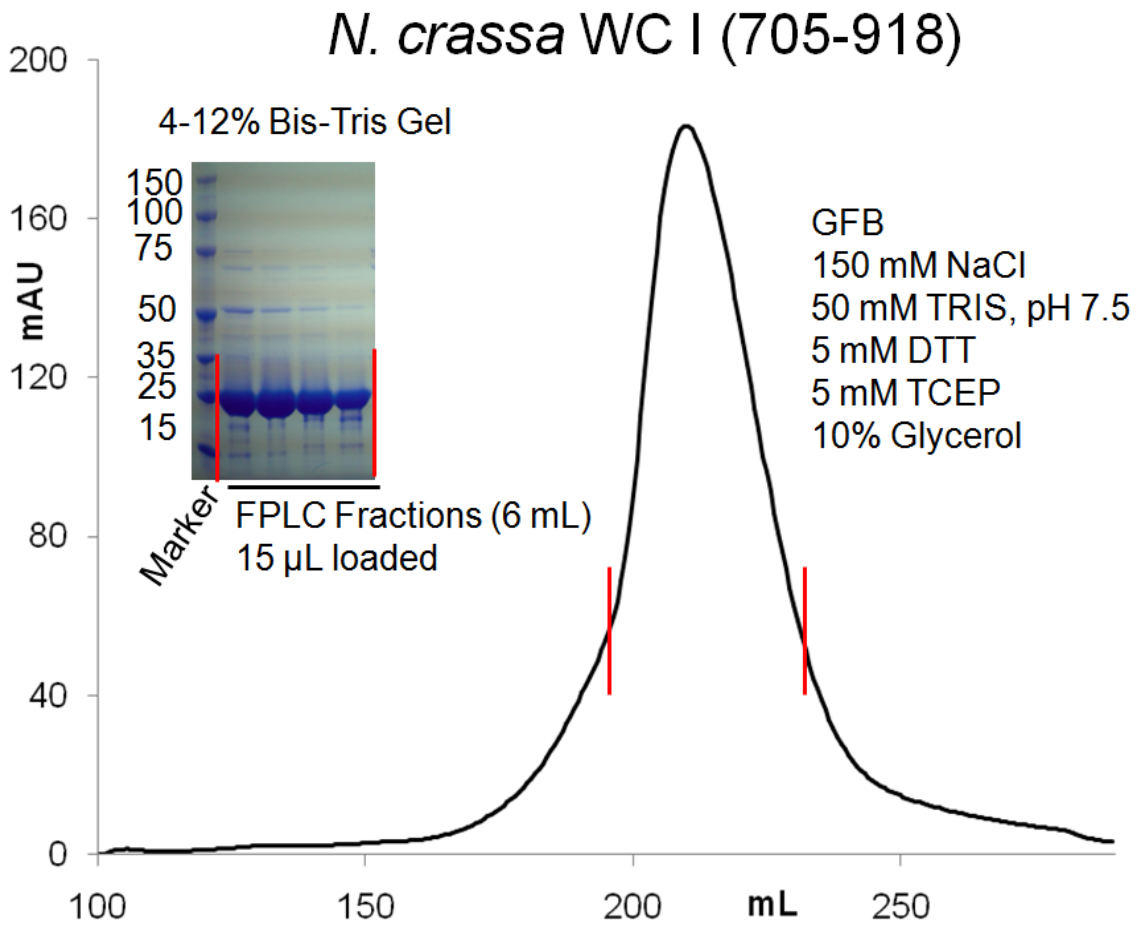

Figure 1.3.2.2 SEC Purification of *N. crassa* WC I (705-918) Shows a Defined Elution Peak. SDS-PAGE gel analysis of SEC elution fractions show soluble protein corresponding to the proteins 25 kDa theoretical molecular weight. The SEC elution profile has a principle elution peak at 210 mL, suggestive of a species with a 38kDa apparent molecular weight, smaller than the predicted 50 kDa molecular weight associated with a dimer. Red bars on SDS-PAGE gel and SEC elution profile indicated protein selected for concentration and crystallization screening.

stoichometric ratio, first using an excess of WC I then WC II at 4, 2, and 1:1. In both cases, the protein at a constant concentration was fixed at 0.6 mg/mL at screened using the hanging drop method (4 µL total drop volume). Conditions were checked daily for three weeks, with the majority of drop conditions for *N.*

crassa WC I (705-918) and WC II (174-455) mixtures either clear drops or light precipitate.

Preparation of *N. crassa* WC I (705-1000) used the same method as described for WC II (174-455), with protein samples taken at 2, 3 and 5 mg/mL during concentration. SDS-PAGE gel analysis of SEC elution fractions (Figure 1.3.2.3) showed a band at 35 kDa, closely matching the proteins theoretical molecular weight of 34 kDa. SDS-PAGE gel analysis also indicated the presence of significant low molecular weight banding, which increase when stored for 16 h at 4 °C (not shown).

N. crassa WC I (705-1000) was screened for crystallization at 0.75, 1.5 and 2.5 mg/mL using the approach taken for N. crassa WC I (705-918). Results from screening showed a transition from clear/light granular precipitate for the lowest concentration to a majority of drops medium granular precipitate for higher concentration.

In an effort to screen *N. crassa* WC I (705-1000) with WC II (179-500) together, the proteins were mixed at 4 °C in an approximately equimolar ratio at 3 mg/mL. Upon mixing the proteins with an identical buffer composition, the solution immediately precipitated. The proteins were then mixed at 1.5 and 1 mg/mL and in each case the protein immediately precipitated. Following precipitation, the samples were centrifuged at 4 °C for 10 m and 13,000 rpm.

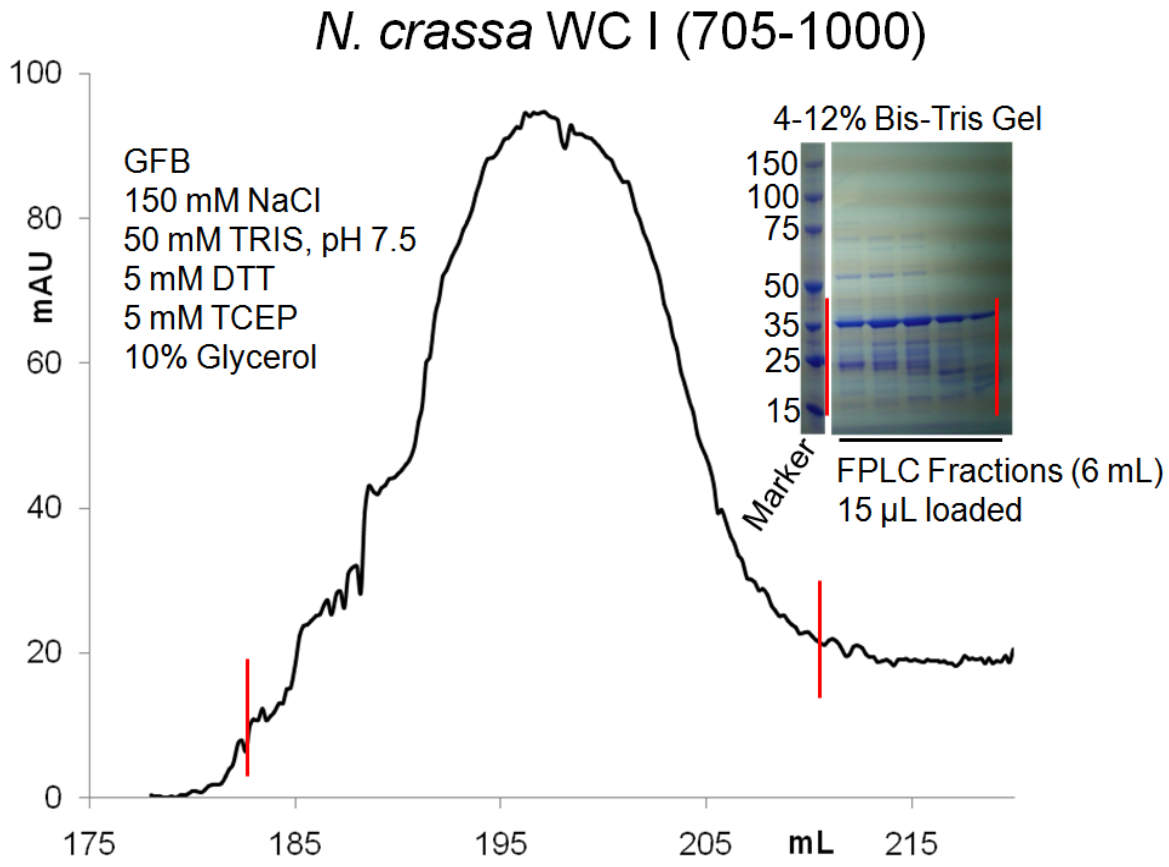

Figure 1.3.2.3 Purification of *N. crassa* WC I (705-1000) Shows Soluble Protein. SDS-PAGE gel analysis of SEC elution fractions show soluble protein corresponding to the theoretical 34 kDa molecular weight. The SEC elution profile indicates a broad elution peak suggestive of multiple oligomeric states in exchange at 198 mL. This elution value is also suggestive a species with an apparent molecular weight of 74 kDa, slightly larger than the predicted 68 kDa theoretical molecular weight for *N. crassa* WC II (705-1000) dimer.

The supernatant was removed and the concentration was tested using the Bradford assay, which detected no protein in the supernatant.

1.3.6 *P. blakesleeanus* MADB Primary Sequence Analysis

To facilitate bioinformatic analysis for *P. blakesleeanus* as described for *N. crassa* WC II, a search of the Pfam and CDD databases using the full-length

primary sequence was first performed. From this search a PAS fold from (62-136) and (43-145) as well as a GATA type zinc finger domain from (313-348) and (312-349) were identified (reported respectively). Additionally, a NLS sequence was observed from (301-306). To evaluate sequence homology, an alignment of *P. blakesleeanus* (numbered) with the studied homologs (Figure 1.3.2.4) was generated. This alignment showed partial and strict residues conservation from Thr30 to Arg50, N-terminal to the identified PAS domain.

From Jpred3 analysis of the primary sequence, secondary structural predictions suggested the presence of an α-helix N-terminal to the PAS domain (26-50). From previous studies, the presence of an α-helix N-terminal to the PAS domain (N-cap), which has been suggested to facilitate stabilizing interactions [73] was predicted, as well as α-helical structures extending from the C-terminal end of the PAS domain to the zinc finger (Figure 1.3.2.5).

XtalPred analysis further predicted α-helices and β-strands both N- and C-terminal to the PAS domain with an α-helical region from (28-48) and disordered regions both N- and C-terminal to the identified PAS fold as well as a coiled coil domain (156-183) and low-complexity region C-terminal to the PAS core (Figure 1.3.2.6).

Based on conclusions drawn from bioinformatic analysis of the *N. crassa* primary sequence and from application of bioinformatic methods to the *P.*

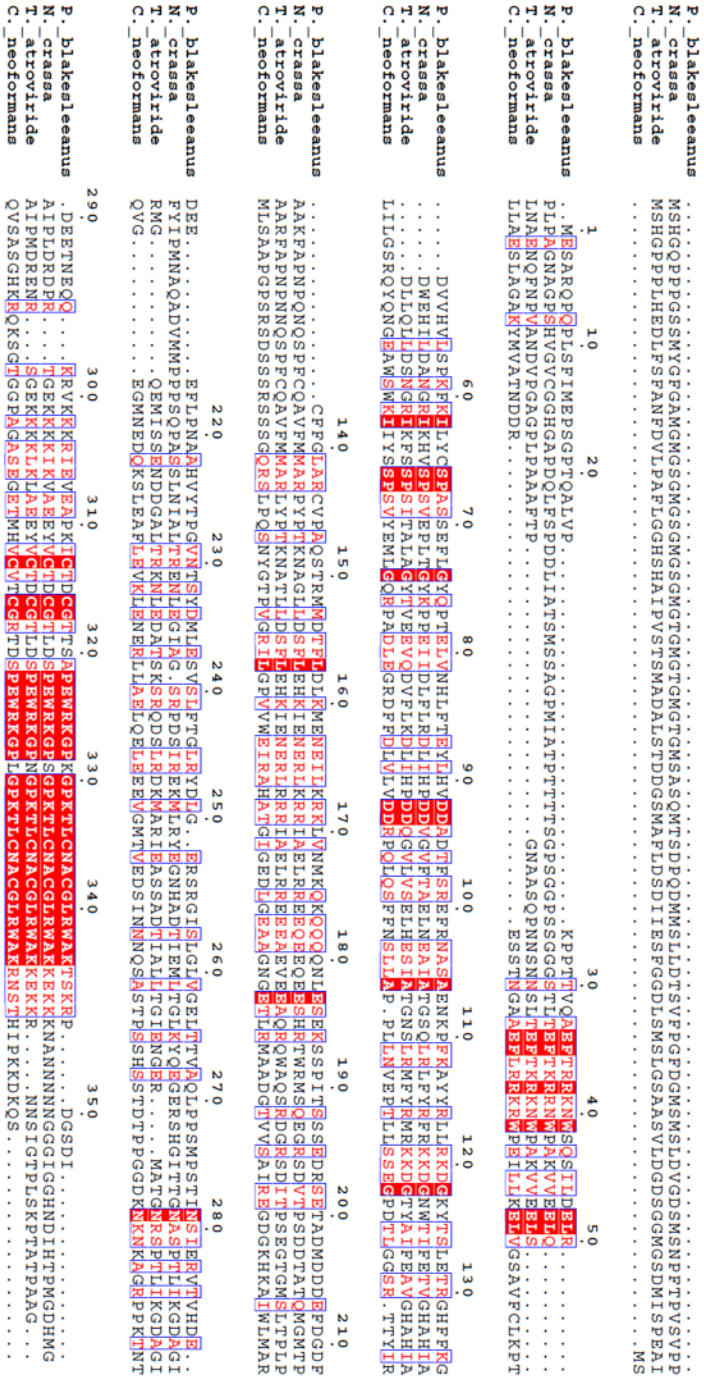

Figure 1.3.2.4 ClustalW Alignment of *P. blakesleeanus* MADB with Studied Homologs. An alignment of full-length *P. blakesleeanus* MADB primary sequence (numbered) with *N. crassa* WC II, *T. atroviride* BLR-2 and *C. neoformans* WC II shows partial (red font) and strict conservation (white font with red highlighting) starting at Glu2 N-terminal to the Pfam identified PAS (62-136) and continuing C-terminally to the zinc finger (313-348).

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80...
MESARQPQPLSFIMEPSGPTQALVPKPPTTVQAEFTRRKNWSQSILDELRDVVHVLSPKFKILYCSPASSEFLGYQPTELVNH
------------------EEEE---HHHHHHHHHHHHHHHHHHHHHHHH-HHEEE-----EEEEHHHHHHH-----HHHH--

.*...90....*..100....*..110....*..120....*..130....*..140....*..150....*..160....*.
LFTEYLHVDDADTFSREFRNASAENKPFKAYYRLLRKDGKYTSLETRGHFFKGCFFGLARCVPAQSTRMMDTFLDLKMENEIL
----------HHHHHHHHHHHH----EEEEEEEE-----EEEEEEEEEEEE---EEEEEEEHHHHHHHHHHHHHHHHHHHH

.170....*..180....*..190....*..200....*..210....*..220....*..230....*..240....*..25
KRKLVNMKQKQQQQNLESEKSSPITSSSEDRSETADMDDDEFDGDFDEEEFLPNAAHVYTPGVNTSYDMLESVSLFTGLRYDL
HHHHHHHHHHHHHHHHHH-----EEEEEE--------------HHHHHHHHHHHHHHHHH---HHE------EEEEE------

0....*..260....*..270....*..280....*..290....*..300....*..310....*..320....*..330..
GERSRGISLGLVGELTTVAQLPPSMPSTINSIERVTVHDEDEETNEQQKRVKKKRIEVEAPKICTDCGTTSAPEWRKGPKGPK
--------HHHH-H--------------------------------HHHHHHHHHHH------EEEE---------------

..*..340....*..350....
TLCNACGLRWAKTSKRPDGSDI
-------HHHHHH---------
```

Figure 1.3.2.5 Jpred3 Secondary Structural Predictions for the *P. blakesleeanus* MADB Primary Sequence. Secondary structural analysis of the full-length *P. blakesleeanus* MADB primary sequence predict α-helices (red) and β-strands (blue) starting at Pro19, N-terminal to the Pfam identified PAS domain (62-136) and continuing to the C-terminal zinc finger (313-348).

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
MESARQPQPLSFIMEPSGPTQALVPKPPTTVQAEFTRRKNWSQSILDELRDVVHVLSPKFKILYCSPASSEFLGYQPTELVNHLFTEYLHVDDADTFSRE

....*..110....*..120....*..130....*..140....*..150....*..160....*..170....*..180....*..190....*..200
FRNASAENKPFKAYYRLLRKDGKYTSLETRGHFFKGCFFGLARCVPAQSTRMMDTFLDLKMENEILKRKLVNMKQKQQQQNLESEKSSPITSSSEDRSET

....*..210....*..220....*..230....*..240....*..250....*..260....*..270....*..280....*..290....*..300
ADMDDDEFDGDFDEEEFLPNAAHVYTPGVNTSYDMLESVSLFTGLRYDLGERSRGISLGLVGELTTVAQLPPSMPSTINSIERVTVHDEDEETNEQQKRV

....*..310....*..320....*..330....*..340....*..350....
KKKRIEVEAPKICTDCGTTSAPEWRKGPKGPKTLCNACGLRWAKTSKRPDGSDI
```

Figure 1.3.2.6 XtalPred Predictions for *P. blakesleeanus* MADB. Secondary structural analysis of the full-length *P. blakesleeanus* MADB primary sequence predict α-helices (red), β-strands (blue), as well as low-complexity (183-195, 202-217) (itializied) and disordered (underlined) regions both N- and C-terminal identified PAS core. Also predicted is presence of a coiled coil (bold) region (156-183) C-terminal to the Pfam identified PAS domain (62-136).

blakesleeanus MADB primary sequence, six PAS domain containing variants were generated (Figure 1.3.2.7) incorporating the predicted N-terminal helix, starting at Pro27 and extending C-terminally, resulting in (27-181, 27-286, 27-289, 27-293, 27-297 and 27-300). An additional variant including the proteins PAS domain and C-terminal zinc finger (27-346), based on homology to *N. crassa* WC II (179-500) was also generated for soluble protein expression.

1.3.6.1 *P. blakesleeanus* MADB Variants

From initial screening efforts, SDS-PAGE gel analysis of small-scale expression tests for each of the seven generated variants failed to show any bands, in either the soluble or insoluble fractions, corresponding to the protein theoretical molecular weight. In an effort to produce *P. blakesleeanus* MADB variants in quantities sufficient for crystallization screening, 25 g preparations were expressed and purified as described for *N. crassa* WC II (179-500). Analysis of IMAC elution fractions from these preparations, using SDS-PAGE, mirrored results from the small-scale tests with protein at or below the limit of detection for the Bradford assay and SDS-PAGE gel as evidence by a lack of detected protein and an absence of bands corresponding to the theoretical molecular weight respectively (not shown). The highest observed soluble protein expression was seen with the shortest PAS variant (27-181); however the protein precipitated during IMAC elution at 4 °C. Supernatant isolated from a centrifugation performed for 10 min at 13,000 rpm was shown to have a final

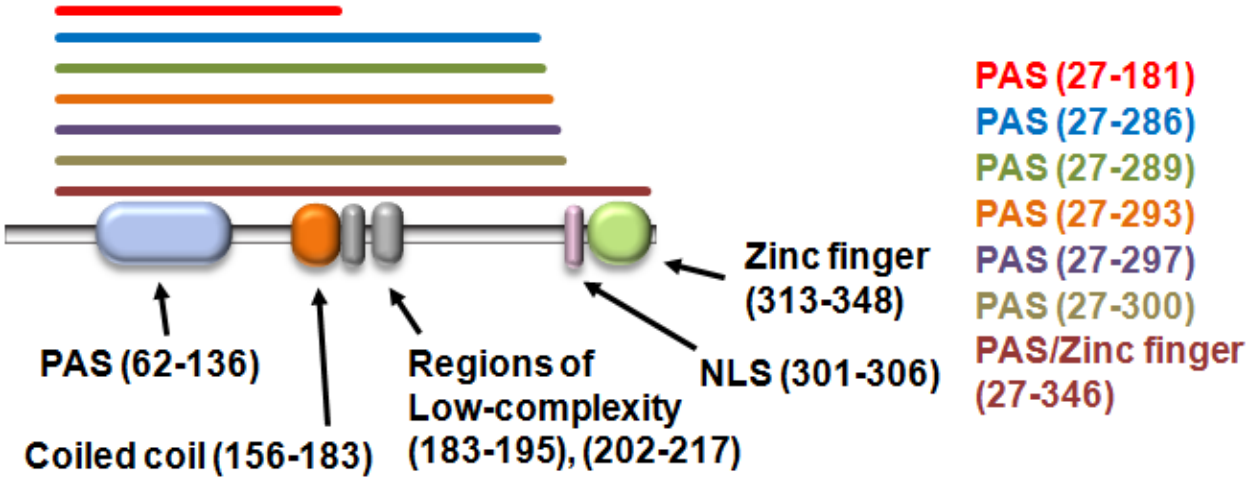

Figure 1.3.2.7 Summary of *P. blakesleeanus* MADB PAS and Zinc finger Domain Containing Variants. Colored bars illustrate inclusion of bioinformatically predicted PAS (62-136), coiled coil (156-183), low-complexity regions. Also shown is *P. blakesleeanus* MADB (27-346), which contains the predicted PAS domain, coiled coil, low-complexity regions, as well as the NLS (301-306) and C-terminal zinc finger (313-348).

concentration of 1.5 mg/mL (not shown).

Crystallization screening was not performed on these preparations.

1.3.7 *T. atroviride* BLR-2 Primary Sequence Analysis

Bioinformatic methods used in the analysis of the *N. crassa* primary sequence were extended to *T. atroviride* for the generation of soluble WC II variants. A search of the proteins primary sequence to Pfam and CDD databases identified a PAS fold from (163-239) and (163-232) as well as a GATA type zinc finger domain from (430-465) and (429-478) respectively. A primary sequence alignment of *T. atroviride* (numbered) with the studied homologs

(Figure 1.3.2.8), showed partial and strict residues conservation N-terminal to the identified PAS domain starting at Asn131 and continuing to Ser151.

Secondary structural analysis with Jpred3 predicted an α-helix (127-150) just N-terminal to the PAS domain with α-helical structures predicted to extend to from the C-terminal end of the PAS domain to Lys420 (Figure 1.3.2.9). XtalPred analysis returned α-helices and β-strands both N- and C-terminal to the PAS domain (Figure 1.3.3) with an α-helical region predicted from (126-149). Predictions of low-complexity and disordered regions were predicted outside of the PAS core with a coiled coil domain predicted for (273-304).

Based on analysis of the *T. atroviride* primary sequence and previous results from expression of *N. crassa* variants, five PAS domain containing fragments were generated incorporating the predicted N-terminal helix starting at Leu133 and extending C-terminally resulting in (133-318, 133-411, 133-415 and 133-417). An additional variant including the proteins C-terminal zinc finger (133-467), based on homology to *N. crassa* WC II (179-500) was also generated for soluble protein expression (Figure 1.3.3.1).

1.3.7.1 *T. atroviride* BLR-2 Variants

Similar to results seen with *N. crassa* WC II (154-285) and (154-323), *T. atroviride* BLR-2 (133-297) SDS-PAGE gel analysis showed no bands in IMAC elution fractions matching the proteins theoretical molecular weight (not shown).

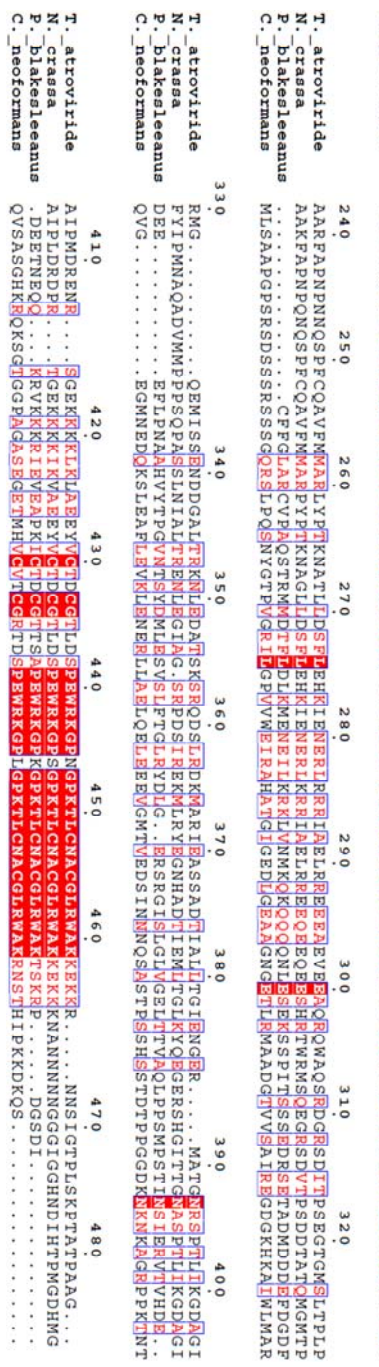

Figure 1.3.2.8 ClustalW Alignment of *T. atroviride* BLR-2 with Studied Homologs. An alignment of the full-length primary sequence for *T. atroviride* BLR-2 (numbered) with full-length *N. crassa* WC II, *P. blakesleeanus* MADB and *C. neoformans* WC II shows partial (red font) and strict conservation (white font with red highlighting) starting at Glu96 N-terminal to the Pfam identified PAS (163-239) and continuing to the end of the C-terminal zinc finger (430-465

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80...
MSHGPPPLHEDLFSFANFDVLPAFLGGHSHAIPVSTSMADALSTDDGSMAFLDSDIIESFGGDLSMSLGSAASVLDGDSGGMG
----------------------HHHHH--------EE------------EEE-------------EEE---HHHH--------

.*...90....*..100....*..110....*..120....*..130....*..140....*..150....*..160....*.
SDMISPEAILNAENQFNPVANDVPGAGPLPAAAFTPGNAASQPNNSNNSLTEFTKRKNWPAKVVEELSDLLQLLDSNGRIKFS
------HHHHHHH------------------------------HHHHHHHHHHHHHHHHHHHHHHHH--EEEEE-----EEEE

.170....*..180....*..190....*..200....*..210....*..220....*..230....*..240....*..25
SPSITALAGYTVEEVQDVFLKDLIHPDDQGVLVSELHESIATGNSLRMFYRMRKKDGTYAIFEAVGHAHIAAARFAPNPNNQS
-HHHHHH-----HHHHH-----------HHHHHHHHHHHHH---EEEEEEEEE-----EEEEEEEEEEEEE------------

0....*..260....*..270....*..280....*..290....*..300....*..310....*..320....*..330..
PFCQAVFMMARLYPTKNATLLDSFLEHKIENERLRRRIAELRREEEAEVEEAQRQWAQSRDGRSDITPSEGTGMSLTPLPRMG
-EEEEEEEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH----------------------------

..*..340....*..350....*..360....*..370....*..380....*..390....*..400....*..410....*
QEMISSENDDGALTRKNLEDATSKSRQDSLRDKMARIEASSADTIALLTGIENGERMATGNRSPTLIKGDAGIAIPMDRENRS
-------------------------------------------HHHHH------------------------------HHHHH

..420....*..430....*..440....*..450....*..460....*..470....*..480....
GEKKKKLKLAEEYVCTDCGTLDSPEWRKGPNGPKTLCNACGLRWAKKEKKRNNSIGTPLSKPTATPAAG
HHHHH-------EEEEE------------------------HHHHHHHHH-------------------
```

Figure 1.3.2.9 Jpred3 Analysis of *T. atroviride* BLR-2 Primary Sequence. Secondary structural predictions for the full-length *T. atroviride* BLR-2 primary sequence predict α-helices (red) and β-strands (blue) starting at Ala23 and continuing to the C-terminal zinc finger (430-465). Also observed is an extended α-helix N- and C-terminal to the Pfam identified PAS domain (163-239).

This contrasted results from the longer PAS variants (133-318, 133-411, 133-415 and 133-417), which showed a higher level of soluble protein expression in purification steps.

SDS-PAGE analysis of IMAC elution fractions for (133-318) showed a principle band at 23 kDa, which matches theoretical molecular weight as well as low molecular weight bands (Figure 1.3.3.2). The SEC profile for protein purified using a HiLoad 26/60 Superdex 75 column, showed a peak eluting at 146 mL suggestive of a species with an apparent molecular weight of 44 kDa, which closely matched the expected value for a 46 kDa dimer (Figure 1.3.3.2). Also

Figure 1.3.3 XtalPred Analysis of *T. atroviride* BLR-2. Secondary structural predictions for the full-length *T. atroviride* BLR-2 primary sequence predict α-helices (red), β-strands (blue), as well as low-complexity (itialized) and disordered (underlined) regions both N- and C-terminal identified PAS core. Analysis also suggests and the presence of a coiled coil (bold) region (273-304) at the C-terminal to the Pfam identified PAS domain (163-239).

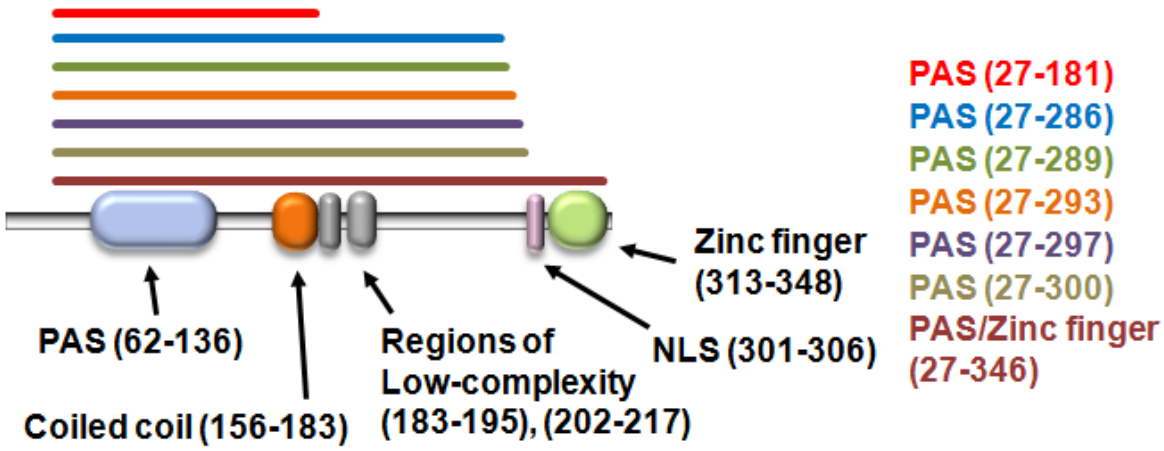

Figure 1.3.3.1 Summary of *T. atroviride* BLR-2 PAS and Zinc finger Domain Containing Variants. Colored bars illustrate inclusion of bioinformatically predicted PAS (163-239), coiled coil (273-304) and low-complexity regions, as well as *T. atroviride* BLR-2 (133-467), which contains the NLS (418-423) and C-terminal zinc finger (430-465).

observed was a second overlapping peak eluting at 163 mL, suggestive of a 23 kDa species.

SDS-PAGE gel analysis of SEC elutions fractions showed a 23 kDa band for both peaks with a gradual increase in low molecular weight banding for fractions containing the 163 mL peak (not shown). Additionally, resolution of these peaks in SEC purifications were unaffected by variations in extraction and gel filtration buffer composition (not shown). As observed in Figure 1.3.3.2, *T. atroviride* BLR-2 (133-318) concentrated to 5 mg/mL showed the low molecular weight bands, reducing the protein purity below levels desired for crystallographic studies.

Consistent with results from (133-318), IMAC and SEC purification for the

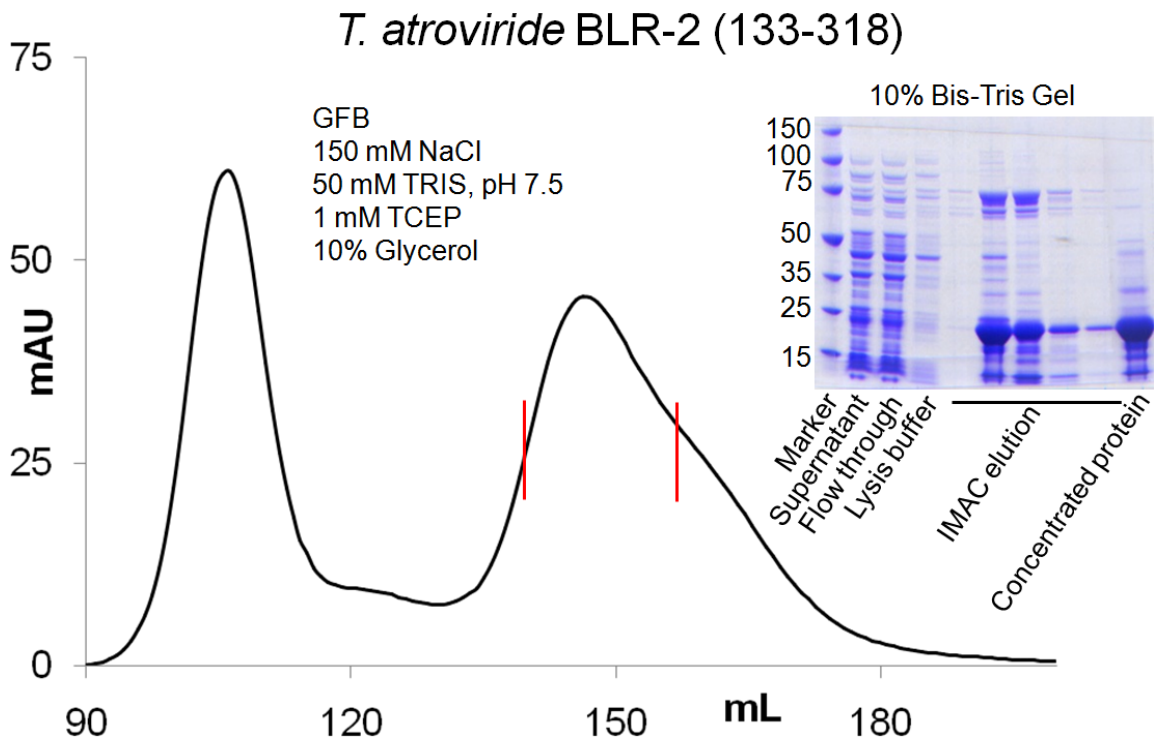

Figure 1.3.3.2 Purification of *T. atroviride* BLR-2 (133-318). SDS-PAGE gel analysis of IMAC elution fractions show bands corresponding to the proteins theoretical 23 kDa molecular weight. SEC purification shows a principle elution peak at 146 mL suggestive of a species with an apparent molecular weight of 44 kDa, which co-elutes with a 163 mL peak (26 kDa apparent molecular weight). Low molecular weight bands are observed from SDS-PAGE analysis of protein concentrated to 5 mg/mL (right). Eluted protein selected for concentration and crystallization screening is denoted with red bars on the SEC elution profile.

longer *T. atroviride* BLR-2 PAS variants (133-411, 133-415 and 133-417) showed low molecular weight bands and a co-eluting peak in elution fractions. Interestingly, the ratio of both relative to the principle elution species was reduced with as increasing variant length (not shown), suggesting the additional residues provided interactions needed for formation of a stable dimer.

As noted, preparation of *T. atroviride* BLR-2 (133-467), containing both the

PAS and zinc finger was preformed. From SDS-PAGE gel analysis of IMAC elution fractions (Figure 1.3.3.3), a band corresponding to the proteins 39 kDa theoretical molecular weight was observed.

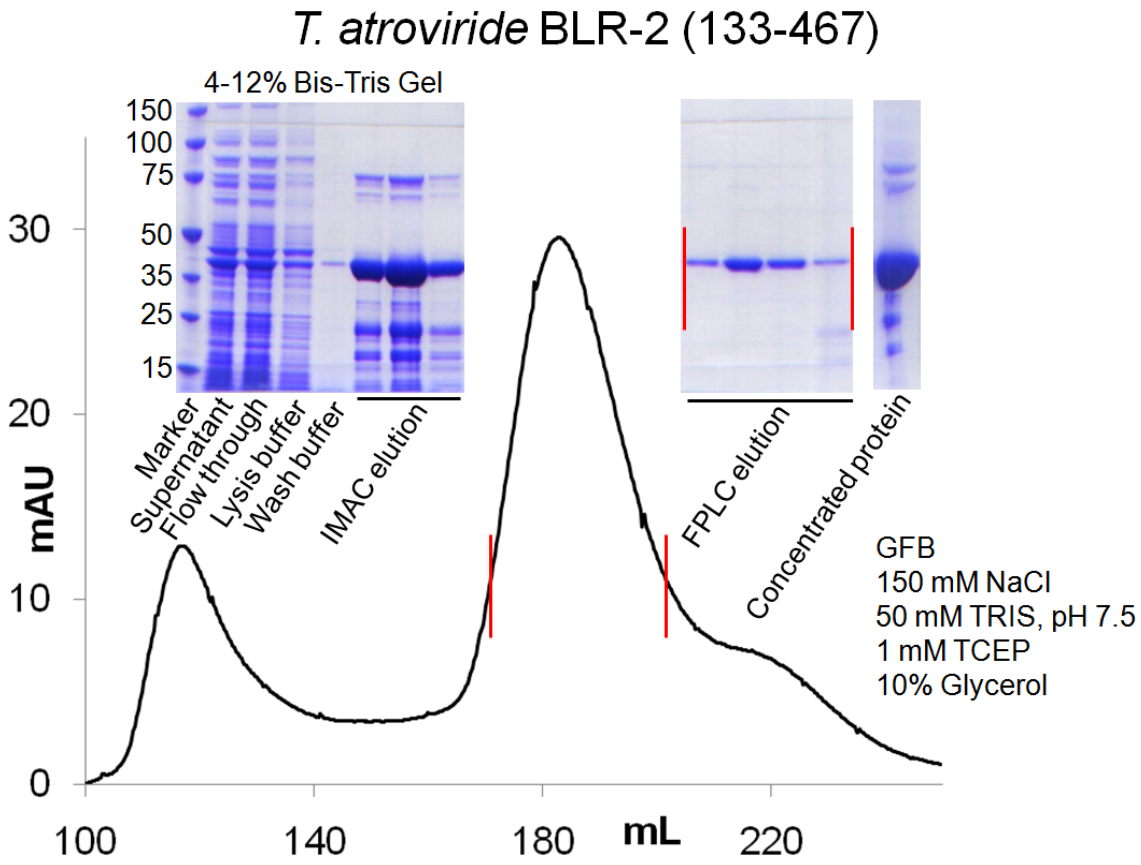

Figure 1.3.3.3 Purification of *T. atroviride* BLR-2 (133-467). SDS-PAGE gel analysis of IMAC elution fractions show bands corresponding to the proteins theoretical 39 kDa molecular weight. A principle SEC elution peak at 186 mL is suggestive of a species with an apparent molecular weight of 108 kDa, larger than the proteins theoretical 78 kDa molecular weight. Red bars on the SEC elution profile and accompanying SDS-PAGE gel denote protein selected for concentration to 5 mg/mL for crystallization screening.

Evaluation of the SEC elution profile showed a principle elution peak at 186 mL, suggestive of a species with an apparent molecular weight of 108 kDa,

larger than the 78 kDa value expected associated with a dimer. Also observed was a decrease in low molecular weight bands relative to the principle band as seen in protein concentrated to 5 mg/mL (Figure 1.3.3.3).

Crystallization screening of *T. atroviride* BLR-2 variants was performed using methods described for *N. crassa* WC II (179-500). Results from these screens showed approximately 90% light precipitate:10% phase separation for PAS domain containing variants. The same precipitate to phase separation ratio was observed for the variants containing the PAS and zinc finger domain, however approximately 30% of the precipitated drops were heavy (dark gray). In all cases, the overall percentage of wells showing heavy precipitate increased with concentration.

1.3.8 *C. neoformans* WC II Primary Sequence Analysis

Application of bioinformatic analysis for generation of *C. neoformans* WC II variants followed the approach taken for *N. crassa* WC II. From a search of the primary sequence using Pfam and CDD, a PAS (74-122 and 70-133) and zinc finger domain (348-383 and 347-383) was identified (reported respectively). As shown in an alignment of *C. neoformans* (numbered) with the homologs studies (Figure 1.3.3.4), sequence homology N-terminal to the identified PAS domain showed partial and strict residues conservation from Asn26 to Val46, N-terminal to the identified PAS domain.

Secondary structural analysis from Jpred3 suggested the presence of an intermittent α-helical structure N-terminal to the identified PAS domain (26-55) with α-helices and β-strands also predicted to extend from the C-terminal end of the PAS domain to Asp282 (Figure 1.3.3.5). XtalPred predicted α-helices and β-strands both N- and C-terminal to the PAS core with an α-helical region from (24-45) (Figure 1.3.3.6). Further, predicted were disordered regions both N- and C-terminal to the PAS fold, a coiled coil domain (254-281) and a discontinuous low-complexity region (259-302).

From bioinformatic analysis of the *C. neoformans* WC II primary sequence, variants starting at Asn26, incorporating the predicted α-helix N-terminal to the identified PAS domain were extended C-terminally to the zinc finger resulting in (26-250, 26-276, 26-282, 26-289, 26-301, 26-306, 26-311, 26-315, 26-317, 26-320, 26-323, 26-329 and 26-340).

Additional variants including the proteins C-terminal zinc finger based on homology to *N. crassa* WC II (179-500) was also generated for soluble protein expression resulting in (26-378, 26-380 and 26-383), as well as full-length protein (1-392) as shown in Figure 1.3.3.7.

1.3.8.1 *C. neoformans* WC II Variants

As seen in the SDS-PAGE gel analysis of *C. neoformans* WC II (26-250), purified using methods described for *N.* crassa WC II (174-455), a band closely

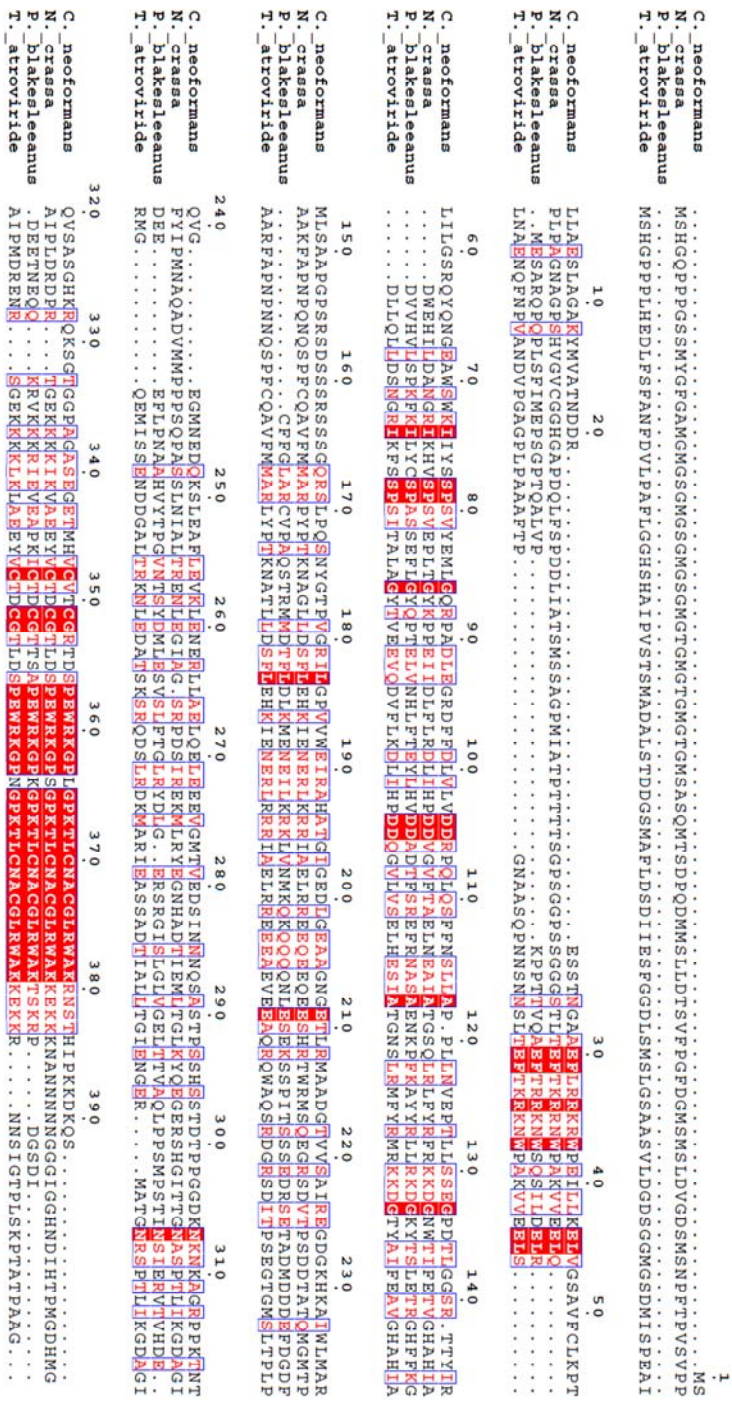

Figure 1.3.3.4 ClustalW Alignment of *C. neoformans* WC II with Studied Homologs. An alignment of the full-length primary sequences for *C. neoformans* WC II (numbered), *N. crassa* WC II, *T. atroviride* BLR-2 and *P. blakesleeanus* MADB shows partial (red font) and strict conservation (white font with red highlighting) starting at Glu9 N-terminal to the Pfam identified PAS (74-122) continuing C-terminally to the zinc finger (348-383).

Figure 1.3.3.5 Jpred3 Analysis of *C. neoformans* WC II Primary Sequence. Secondary structural predictions for the full-length *C. neoformans* WC II primary sequence predict α-helices (red) and β-strands (blue) starting at Met1 and continuing to the C-terminal zinc finger.

matching the proteins 26 kDa theoretical molecular weight was observed (Figure 1.3.3.8). Additionally, protein concentrated to 67 mg/mL, from collected SEC elution fractions showed purity sufficient for crystallization screening as SDS-PAGE gels analysis as indicated minimal high and low molecular weight band formation.

The SEC elution profile for (26-250) showed a principle peak at 229 mL suggestive of a species with an apparent 29 kDa molecular weight, close to the value expected for a monomer. This result contrasted the observation that the SEC elution profiles for each of the longer *C. neoformans* WC II variants all displayed elution peaks consistent with formation of an apparent dimer (not shown).

Figure 1.3.3.6 XtalPred Analysis of *C. neoformans* WC II. Secondary structural predictions for the full-length *C. neoformans* WC II primary sequence predict α-helices (red), β-strands (blue), as well as discontinuous low-complexity (259-302) (itializied) and disordered (underlined) regions both N- and C-terminal identified PAS core. Analysis also suggests and the presence of a coiled coil (bold) region (254-281) at the C-terminal to the Pfam identified PAS domain (74-122).

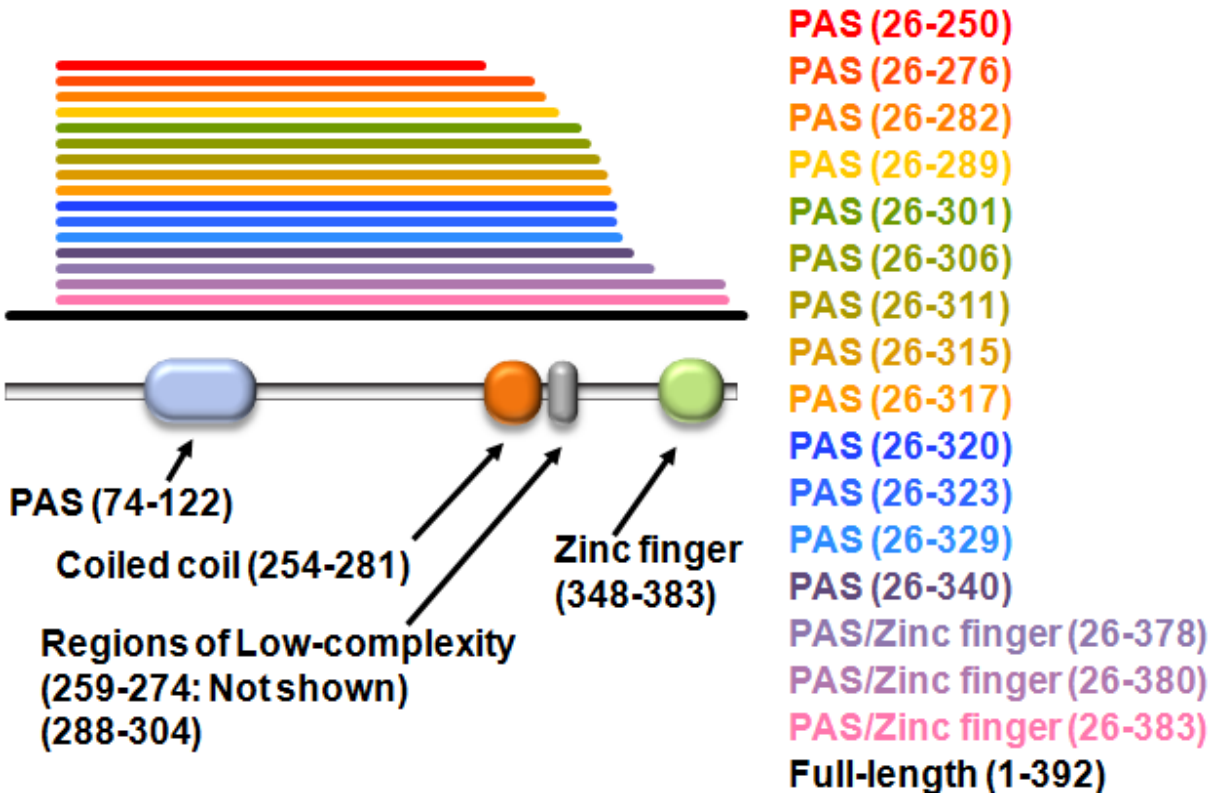

Figure 1.3.3.7 Summary of *C. neoformans* WC II PAS and Zinc finger Domain Containing Variants. Colored bars illustrate variants that either include or exclude the bioinformatically predicted PAS (74-122), coiled coil (254-281), low-complexity regions and C-terminal zinc finger (348-383).

Based on results obtained from purification of the *C. neoformans* WC II variants, the full-length protein (1-392) was generated for study. As seen in Figure 1.3.3.9, bands corresponding to the proteins 45 kDa theoretical molecular weight are observed in IMAC and SEC elutions. Similar to results from the truncated variants, the protein shows relatively little formation of high and low molecular weight bands in SDS-PAGE gel analysis when purified using methods described for *N. crassa* WC II (174-455).

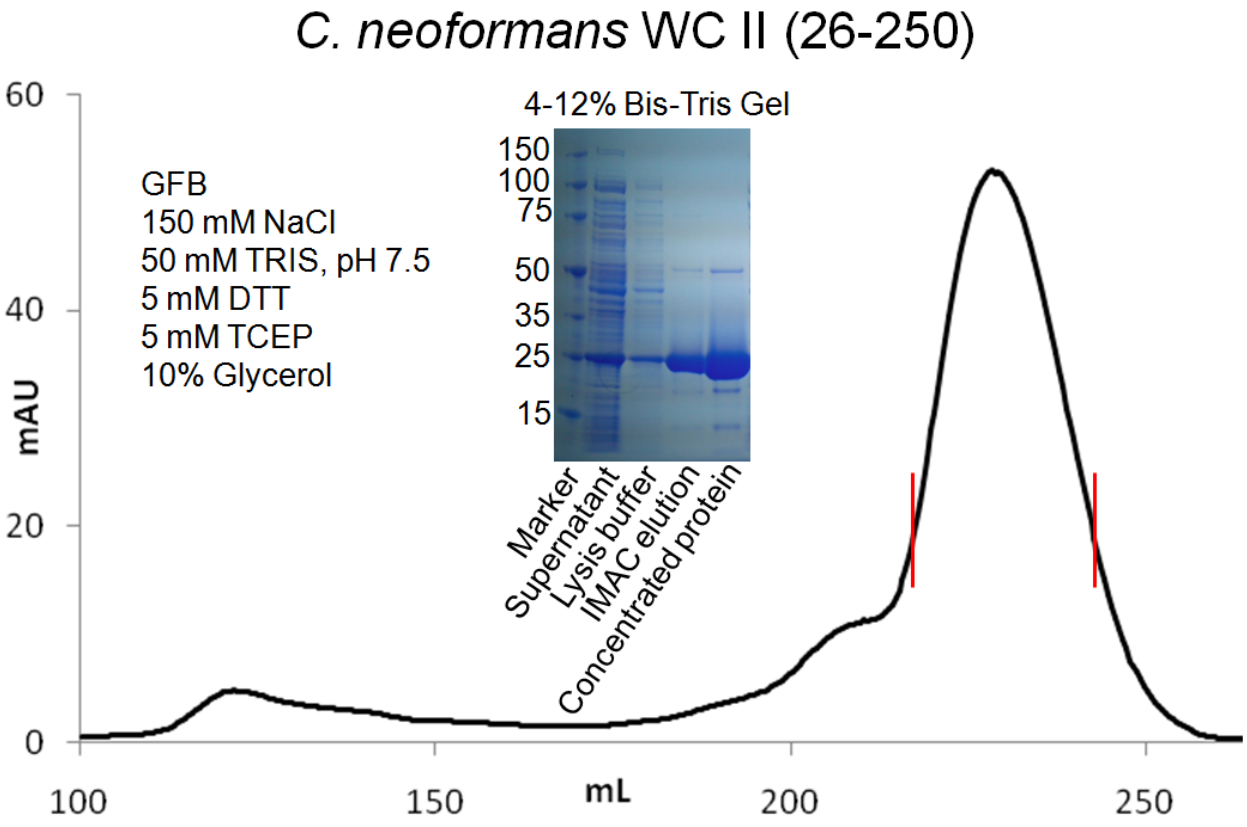

Figure 1.3.3.8 Purification of *C. neoformans* WC II (26-250). SDS-PAGE gel analysis of the IMAC elution shows a band corresponding to the proteins theoretical 26 kDa molecular weight. From SEC purification, the protein was observed to elute at 229 mL, suggestive of a 29 kDa monomeric species. SEC elutions selected for and subsequently concentrated to 67 mg/L for crystallization screening is denoted with red bars on the SEC elution profile.

Crystallization screening of *C. neoformans* WC II proteins was performed using methods described for *N. crassa* WC II (179-500). Results from these screens showed approximately 30-40% phase separation accompanying an equal distribution of clear and precipitated drops for PAS domain containing variants. An approximate ratio of 70% phase separation to clear and precipitated drops was observed for the variants containing the PAS and zinc finger domain.

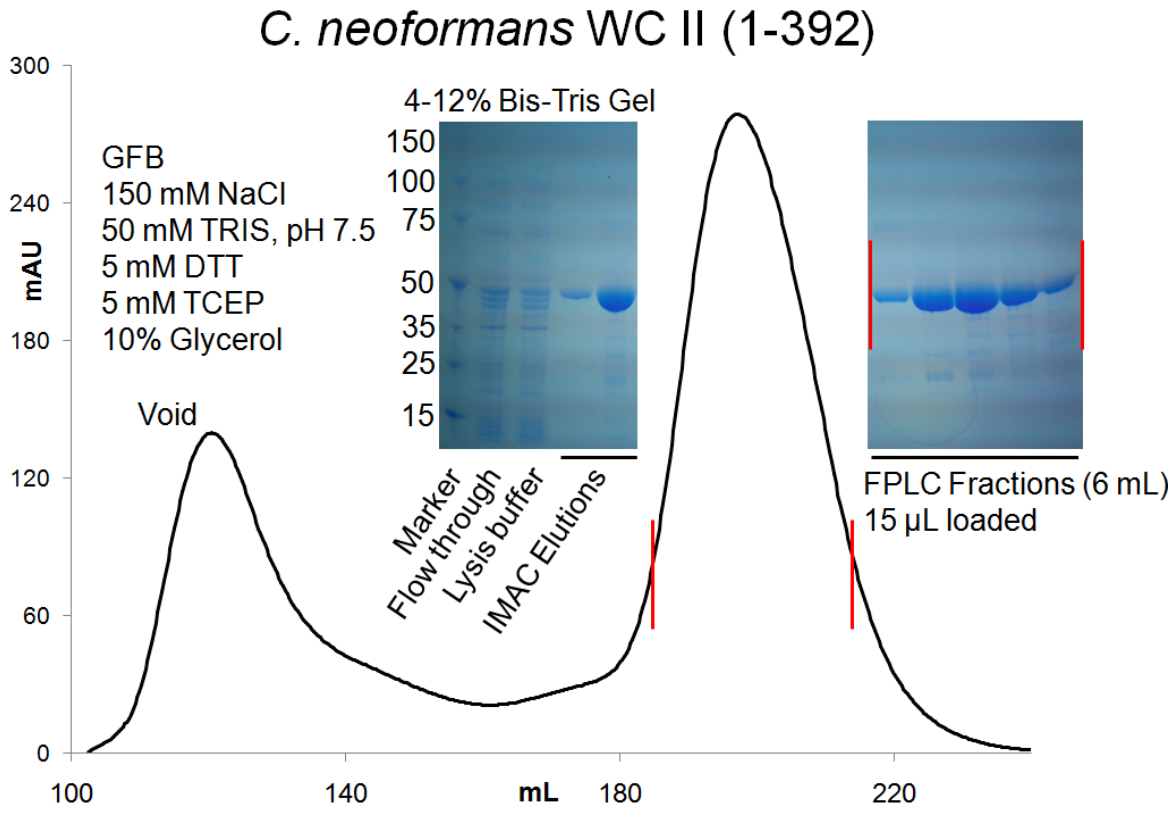

Figure 1.3.3.9 Purification of *C. neoformans* WC II (1-392). SDS-PAGE gel analysis of IMAC and SEC elution fractions show a band corresponding to the proteins 45 kDa theoretical molecular weight. A principle SEC elution peak at 194 mL is suggestive of a species with an apparent molecular weight of 90 kDa, suggestive of a dimer. Eluted protein selected for concentration and crystallization screening are denoted with red bars on the SEC elution profile and corresponding SDS-PAGE gel.

1.3.9 *C. neoformans* WC I Variants

Design of *C. neoformans* White collar I variants was based on a 36 amino acid N-terminally truncated *N. crassa* VVD (VVD-36) fragment previously crystallized by coworkers [74]. From alignment of VVD-36, *N. crassa* WC I and *C. neoformans* WC I LOV domains, His474 in the *C. neoformans* primary sequence is shown to correspond to the start of VVD-36 (Figure1.3.4).

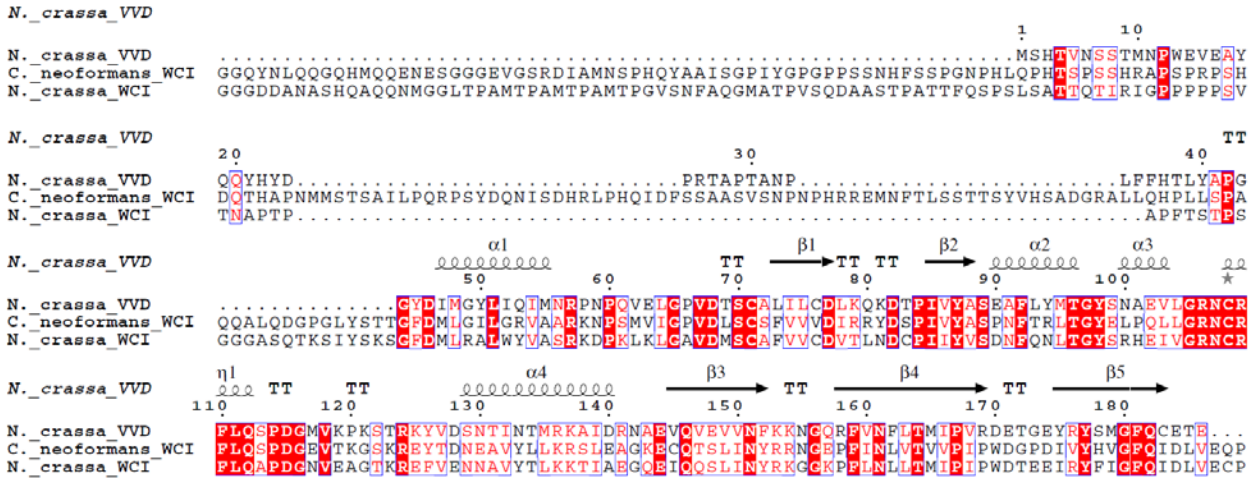

Figure 1.3.4 ClustalW Alignment of LOV Domains from *N. crassa* VVD, WC I and *C. neoformans* WCI Show Residue Conservation. An alignment of the full-length primary sequence for *N. crassa* Vivid (VVD) (numbered), *N. crassa* WC I and *C. neoformans* WC I shows partial (red font) and strict conservation (white font with red highlighting) for the proteins Light Oxygen Voltage (LOV) domain. From the alignment *C. neoformans* His474 is shown to correspond to His37 in the VVD sequence. The secondary structure shown above the sequence alignment, which denotes α-helices, β-strands and turns as determined from the deposited N-terminally truncated VVD-36 crystal structure (PDB: 2PD7).

To generate additional variants, Ile456 and Pro431, which both correspond to Asp24 in the VVD-36 sequence, were also made. In each case, the variants were C-terminally truncated at Val637, which corresponds to the end of the VVD-36 sequence.

For each of the designed variants, 25 gram preparations were expressed and purified using conditions detailed for VVD-36. From SDS-PAGE gel analysis of IMAC elution fractions, no bands corresponding to the variants theoretical molecular weight were observed (not shown).

1.4 Discussion

To facilitate the study of White Collar transcriptional activators, variants based on bioinformatic analysis of the proteins domain architecture were selected for expression and purification. From these studies, fragments containing the PAS and zinc finger domains were generated in quantities and purity sufficient for biophysical characterization as discussed in Chapter 2.

While results from the bioinformatic methods used were similar to those previously described in identification of the proteins PAS, coiled coil, low-complexity and zinc finger regions [8, 37, 39, 62, 71, 75], it was also observed that extended variants, which include conserved residues C-terminal to the identified PAS domains, demonstrated greater solubility and eluted with apparent volumes suggestive of apparent dimers in SEC purifications.

With these observations in mind, *N. crassa* WC I and II sequences were compared to the known crystal structures of the *M. musculus* heterodimeric CLOCK:BMAL1 transcriptional activator complex [10], as well as the homodimeric negative clock elements Period 2 homolog from *M. musculus* [76] and Period from *D. melanogaster* [35]. To facilitate this alignment, the WC I and II sequences were first compared with the canonical PAS A core (Aβ, Bβ, Cα, Dα, Eα, Fα, Gβ, Hβ and Iβ) [10], followed by the PAS B core (Figure 1.4.1). From these alignments, WC I (586-705) and (706-806) showed strict and partial

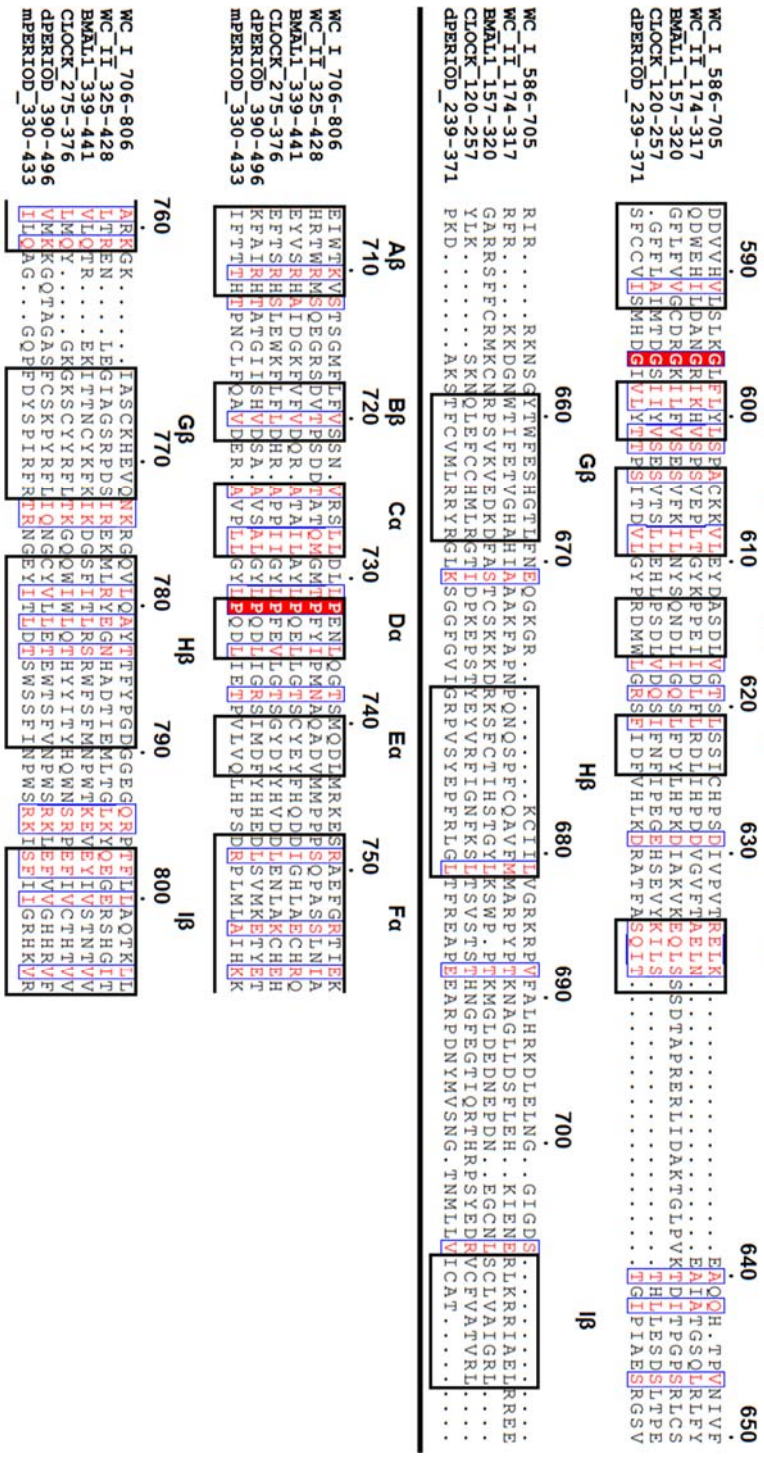

Figure 1.4.1 ClustalW Alignment of *N. crassa* WC I and II Suggest the Presence a PAS Repeat. Alignment of *N. crassa* WC I (numbered) and II with *M. musculus* CLOCK:BMAL1 (PDB: 4F3L), Period2 (mPERIOD)(PDB: 3GD1) and *D. melanogaster* Period (dPERIOD)(PDB:3RTY) show partial (red font) and strict conservation (white font with red highlighting) with each of the proteins core PAS regions.

conservation with the known structures PAS A and B cores respectively (Figure 1.4.2) [10].

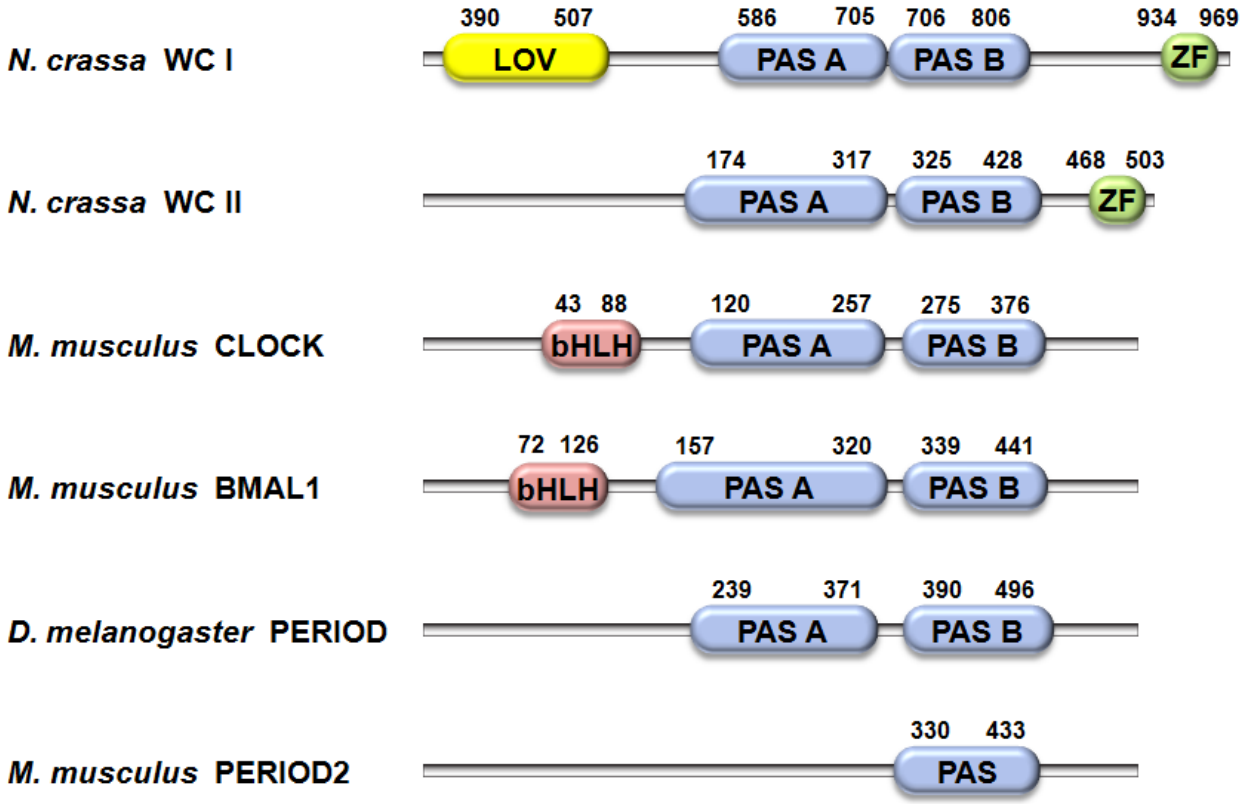

Figure 1.4.2 PAS AB Domain Repeat is Observed in *N. crassa* White Collar I and II. From CLUSTALW sequence alignments, residue homology between *N. crassa* WC I and II proteins to *M. musculus* CLOCK, BMAL1, PERIOD2 homolog and *D. melanogaster* PERIOD PAS core regions suggest the presence of a PAS AB repeat.

Interestingly, while previous analysis of the WC II sequence did not predict the presence of a PAS AB repeat, alignment of (174-317) to the known structures respective PAS A cores shows good sequence conservation for N-terminal portion (Aβ, Bβ, Cα, Dα, Eα and Fα), with little homology for the C-terminal region (Gβ, Hβ and Iβ). This alignment compliments the alignment of (325-441)

to the PAS B regions, which shows good agreement throughout the entire PAS core (Figure 1.4.1). Results from these alignments suggest the presence of a PAS AB repeat in both *N. crassa* WC I and II similar to the arrangement observed in *M. musculus* CLOCK:BMAL1 and Period proteins from *D. melanogaster* and *M. musculus* (Figure 1.4.2).

Based on the results from WC I and II protein variant screening as well as bioinformatic analysis of the primary sequence, which suggest the presence of a WC I and II PAS AB repeat, further efforts to purify both WC I and II based on the putative PAS AB repeats, offers the possibility of furthering our understanding of how these proteins associate using biophysical methods such as X-ray crystallography or small-angle X-ray scattering.

To facilitate these studies, WC I PAS domain containing efforts should be centered on (585-806), with N- and C-terminal variations that might include flanking regions needed for crystallization. For WC II, PAS only variants should focus on (154-445) and (174-445), which extend to the C-terminal end of the putative PAS AB repeat and includes or excludes the N-terminal α-helix (putative N-cap) respectively.

1.5 References

1. Brunner, M. and T. Schafmeier Transcriptional and post-transcriptional regulation of the circadian clock of cyanobacteria and Neurospora. *Genes Dev*. (2006). 20(9): 1061-74.

2. Dunlap, J.C. Molecular bases for circadian clocks. *Cell*. (1999). 96(2): 271-90.

3. Heintzen, C. and Y. Liu The Neurospora crassa circadian clock. *Adv Genet*. (2007). 58: 25-66.

4. Tataroglu, O. and T. Schafmeier Of switches and hourglasses: regulation of subcellular traffic in circadian clocks by phosphorylation. *EMBO Rep*. (2010). 11(12): 927-35.

5. Loros, J.J. and J.C. Dunlap Genetic and molecular analysis of circadian rhythms in Neurospora. *Annu Rev Physiol*. (2001). 63: 757-94.

6. Jinhu, G. and L. Yi Molecular mechanism of the Neurospora circadian oscillator. *Protein Cell*. (2010). 1(4): 331-41.

7. Lee, K., J.C. Dunlap, and J.J. Loros Roles for WHITE COLLAR-1 in circadian and general photoperception in Neurospora crassa. *Genetics*. (2003). 163(1): 103-14.

8. Linden, H. and G. Macino White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in Neurospora crassa. *EMBO J*. (1997). 16(1): 98-109.

9. Denault, D.L., J.J. Loros, and J.C. Dunlap WC-2 mediates WC-1-FRQ interaction within the PAS protein-linked circadian feedback loop of Neurospora. *EMBO J*. (2001). 20(1-2): 109-17.

10. Huang, N., et al. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science*. (2012). 337(6091): 189-94.

11. Moglich, A., R.A. Ayers, and K. Moffat Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure*. (2009). 17(10): 1282-94.

12. Taylor, B.L. and I.B. Zhulin PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev*. (1999). 63(2): 479-506.

13. Amezcua, C.A., et al. Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation. *Structure*. (2002). 10(10): 1349-61.

14. Cheng, P., et al. PAS domain-mediated WC-1/WC-2 interaction is essential for maintaining the steady-state level of WC-1 and the function of both proteins in circadian clock and light responses of Neurospora. *Mol Cell Biol*. (2002). 22(2): 517-24.

15. Talora, C., et al. Role of a white collar-1-white collar-2 complex in blue-light signal transduction. *EMBO J*. (1999). 18(18): 4961-8.

16. Crosson, S., S. Rajagopal, and K. Moffat The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. *Biochemistry*. (2003). 42(1): 2-10.

17. Moglich, A. and K. Moffat Structural basis for light-dependent signaling in the dimeric LOV domain of the photosensor YtvA. *J Mol Biol*. (2007). 373(1): 112-26.

18. Froehlich, A.C., et al. White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science*. (2002). 297(5582): 815-9.

19. Cheng, P., et al. Coiled-coil domain-mediated FRQ-FRQ interaction is essential for its circadian clock function in Neurospora. *EMBO J*. (2001). 20(1-2): 101-8.

20. Dunlap, J.C. and J.J. Loros How fungi keep time: circadian system in Neurospora and other fungi. *Curr Opin Microbiol*. (2006). 9(6): 579-87.

21. Shi, M., et al. FRQ-interacting RNA helicase mediates negative and positive feedback in the Neurospora circadian clock. *Genetics*. (2010). 184(2): 351-61.

22. Cha, J., et al. Control of WHITE COLLAR localization by phosphorylation is a critical step in the circadian negative feedback process. *EMBO J*. (2008). 27(24): 3246-55.

23. Dunlap, J.C. Proteins in the Neurospora circadian clockworks. *J Biol Chem*. (2006). 281(39): 28489-93.

24. Querfurth, C., et al. Posttranslational regulation of Neurospora circadian clock by CK1a-dependent phosphorylation. *Cold Spring Harb Symp Quant Biol*. (2007). 72: 177-83.

25. Tataroglu, O., et al. Glycogen synthase kinase is a regulator of the circadian clock of Neurospora crassa. *J Biol Chem*. (2012). 287(44): 36936-43.

26. Baker, C.L., J.J. Loros, and J.C. Dunlap The circadian clock of Neurospora crassa. *FEMS Microbiol Rev*. (2012). 36(1): 95-110.

27. Liu, Y. and D. Bell-Pedersen Circadian rhythms in Neurospora crassa and other filamentous fungi. *Eukaryot Cell*. (2006). 5(8): 1184-93.

28. Cheng, P., et al. WHITE COLLAR-1, a multifunctional neurospora protein involved in the circadian feedback loops, light sensing, and transcription repression of wc-2. *J Biol Chem*. (2003). 278(6): 3801-8.

29. Heintzen, C., J.J. Loros, and J.C. Dunlap The PAS protein VIVID defines a clock-associated feedback loop that represses light input, modulates gating, and regulates clock resetting. *Cell*. (2001). 104(3): 453-64.

30. Schafmeier, T. and A.C. Diernfellner Light input and processing in the circadian clock of Neurospora. *FEBS Lett*. (2011). 585(10): 1467-73.

31. Hamilton, E.E. and S.A. Kay SnapShot: circadian clock proteins. *Cell*. (2008). 135(2): 368-368 e1.

32. Brown, S.A., E. Kowalska, and R. Dallmann (Re)inventing the circadian feedback loop. *Dev Cell*. (2012). 22(3): 477-87.

33. Zoltowski, B.D., et al. Conformational switching in the fungal light sensor Vivid. *Science*. (2007). 316(5827): 1054-7.

34. Zoltowski, B.D., et al. Structure of full-length Drosophila cryptochrome. *Nature*. (2011). 480(7377): 396-9.

35. Yildiz, O., et al. Crystal structure and interactions of the PAS repeat region of the Drosophila clock protein PERIOD. *Mol Cell*. (2005). 17(1): 69-82.

36. Vreede, J., et al. PAS domains. Common structure and common flexibility. *J Biol Chem*. (2003). 278(20): 18434-9.

37. Ballario, P., et al. White collar-1, a central regulator of blue light responses in Neurospora, is a zinc finger protein. *EMBO J*. (1996). 15(7): 1650-7.

38. Sanz, C., et al. Phycomyces MADB interacts with MADA to form the primary photoreceptor complex for fungal phototropism. *Proc Natl Acad Sci U S A*. (2009). 106(17): 7095-100.

39. Idnurm, A., et al. The Phycomyces madA gene encodes a blue-light photoreceptor for phototropism and other light responses. *Proc Natl Acad Sci U S A*. (2006). 103(12): 4546-51.

40. Casas-Flores, S., et al. BLR-1 and BLR-2, key regulatory elements of photoconidiation and mycelial growth in Trichoderma atroviride. *Microbiology*. (2004). 150(Pt 11): 3561-9.

41. Lu, Y.K., K.H. Sun, and W.C. Shen Blue light negatively regulates the sexual filamentation via the Cwc1 and Cwc2 proteins in Cryptococcus neoformans. *Mol Microbiol*. (2005). 56(2): 480-91.

42. Bernstein, F.C., et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*. (1977). 112(3): 535-42.

43. Sayers, E.W., et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. (2009). 37(Database issue): D5-15.

44. Altschul, S.F., et al. Basic local alignment search tool. *J Mol Biol*. (1990). 215(3): 403-10.

45. Altschul, S.F., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. (1997). 25(17): 3389-402.

46. Finn, R.D., et al. The Pfam protein families database. *Nucleic Acids Res*. (2010). 38(Database issue): D211-22.

47. Marchler-Bauer, A., et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*. (2011). 39(Database issue): D225-9.

48. Cole, C., J.D. Barber, and G.J. Barton The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*. (2008). 36(Web Server issue): W197-201.

49. Slabinski, L., et al. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics*. (2007). 23(24): 3403-5.

50. Slabinski, L., et al. The challenge of protein structure determination--lessons from structural genomics. *Protein Sci*. (2007). 16(11): 2472-82.

51. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. (1999). 292(2): 195-202.

52. Wootton, J.C. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem*. (1994). 18(3): 269-85.

53. Thompson, J.D., D.G. Higgins, and T.J. Gibson CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. (1994). 22(22): 4673-80.

54. Gouet, P., et al. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*. (1999). 15(4): 305-8.

55. Halford, S.E., et al. MSI-low, a real phenomenon which varies in frequency among cancer types. *J Pathol*. (2003). 201(3): 389-94.

56. Famous, K.R., et al. Phosphorylation-dependent trafficking of GluR2-containing AMPA receptors in the nucleus accumbens plays a critical role in the reinstatement of cocaine seeking. *J Neurosci*. (2008). 28(43): 11061-70.

57. Gorynia, S., et al. Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J Struct Biol*. (2011). 176(3): 279-91.

58. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. (1976). 72: 248-54.

59. Nardini, M., et al. The C-terminal domain of the transcriptional corepressor CtBP is intrinsically unstructured. *Protein Sci*. (2006). 15(5): 1042-50.

60. Waldo, G.S., et al. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* (1999). 17(7): 691-695.

61. Nambu, J.R., et al. The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell*. (1991). 67(6): 1157-67.

62. Ballario, P., et al. Roles in dimerization and blue light photoresponse of the PAS and LOV domains of Neurospora crassa white collar proteins. *Mol. Microbiol.* (1998). 29(3): 719-29.

63. Hughes, S.H. and A.M. Stock, *International Tables for Crystallography (2006): Volume F, Crystallography of biological macromolecules.*, in *3.1. Preparing recombinant proteins for X-ray crystallography.*, C.W. Carter and J.J. Müller, Editors. 2006, John Wiley & Sons, Ltd. p. 65-80.

64. Vaidya, A.T., et al. Structure of a light-activated LOV protein dimer that regulates transcription. *Sci Signal*. (2011). 4(184): ra50.

65. Lee, B. and D. Lee Protein comparison at the domain architecture level. *BMC Bioinformatics*. (2009). 10 Suppl 15: S5.

66. Blundell, T. and M.J.E. Sternberg Computer-aided design in protein engineering. *Trends Biotechnol.* (1985). 3(9): 228-235.

67. Zhang, X.C., et al. Evolutionary dynamics of protein domain architecture in plants. *BMC Evol Biol*. (2012). 12: 6.

68. Cuff, J.A. and G.J. Barton Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*. (2000). 40(3): 502-11.

69. Ward, J.J., et al. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. (2004). 20(13): 2138-9.

70. Chen, L., et al. TargetDB: a target registration database for structural genomics projects. *Bioinformatics*. (2004). 20(16): 2860-2.

71. Moriwaki, A., et al. Cloning and characterization of the BLR2, the homologue of the blue-light regulator of Neurospora crassa WC-2, in the phytopathogenic fungus Bipolaris oryzae. *Curr Microbiol*. (2008). 56(2): 115-21.

72. Lee, J., et al. Changes at the KinA PAS-A dimerization interface influence histidine kinase function. *Biochemistry*. (2008). 47(13): 4051-64.

73. Watts, K.J., et al. Function of the N-terminal cap of the PAS domain in signaling by the aerotaxis receptor Aer. *J Bacteriol*. (2006). 188(6): 2154-62.

74. Zoltowski, B.D. and B.R. Crane Light activation of the LOV protein vivid generates a rapidly exchanging dimer. *Biochemistry*. (2008). 47(27): 7012-9.

75. Liu, Y., Q. He, and P. Cheng Photoreception in Neurospora: a tale of two White Collar proteins. *Cell Mol Life Sci*. (2003). 60(10): 2131-8.

76. Hennig, S., et al. Structural and functional analyses of PAS domain interactions of the clock proteins Drosophila PERIOD and mouse PERIOD2. *PLoS Biol*. (2009). 7(4): e94.

CHAPTER 2

CHARACTERIZATION OF A *NEUROSPORA CRASSA* WHITE COLLAR PAS DOMAIN MEDIATED COMPLEX

2.1 Introduction

As discussed in Chapter 1, circadian clocks are a common feature among biological systems that allow organisms to anticipate diurnal changes in their environment. Central to the *N. crassa* clock are the White Collar I [1] and II [2] (WC I and WC II) transcriptional activators [3], that associate via PAS domains [4] to form a heteromeric complex called the White Collar Complex (WCC) [5]. Once formed, imperfect GATA-type zinc fingers [5, 6] in the WC proteins bind the pseudo-palandromic sequences identified from mutational analysis of the *frq* promoter region (Clock-box) and promote expression of the negative element Frequency [7, 8].

While considerable work has been done to elucidate the role of the WCC in blue-light photoreception and transcription of clock controlled genes as well as negative clock elements, relatively little is known about the proteins on the molecular level, including the oligomeric state of the WCC, orientation of the PAS and zinc finger domains in the complex and how it interacts with the negative clock elements Vivid and Frequency.

With this in mind, the individual transcriptional activators WC I and II from *N.* crassa, as described in Chapter 1, were copurified allowing for characterization of variants containing only the proteins the PAS as well as PAS with the C-terminal zinc finger domain using small-angle X-ray scattering. From these studies, we show the formation of a WC I/II PAS mediated complex, suggesting the *N. crassa* WCC may assemble to form a heterotetermeric species. Formation of this species would allow for a single transcriptional complex to simultaneously bind four cis-acting pseudo-palandromic sequences light-acting response elements identified in the *frq* promoter region [9].

2.2 Materials and Methods

2.2.1 Generating WC I and II Proteins for Small-angle X-ray Scattering Studies

Expression and purification methods for generation of *N. crassa* White Collar I C-terminal Per-ARNT-Sim (PAS C) domain [4, 10, 11], amino acid residue 705-918 (705-918), PAS C with the C-terminal zinc finger (705-1000) [10], as well as *N. crassa* WC II PAS (174-455) and PAS with the C-terminal zinc finger (179-500) variants for small-angle X-ray scattering studies are described in Chapter 1.

2.2.1.1 Purifying WC I/II Proteins

Approximately equal quantities of each frozen cell pellet were first resuspended and mixed on ice with an approximate ratio of 10 mL lysis buffer containing 300 mM NaCl, 50 mM Tris pH 7.5, 10 mM imidazole, 10% v/v glycerol, 2 mM dithiothreitol (DTT) and 5 mM Tris(2-carboxyethyl)phosphine (TCEP) to each gram of cellular pellet. The suspension was then mixed and lysed on ice with four alternating 2 s on/2 s off sonication cycles each lasting 30 s. All purification and handling steps from this point were conducted as rapidly as possible at 4 °C. Following sonication, the proteins were centrifuged at 4 °C for 45 min at 22,000 rpm and the supernatant was first collected, then applied to an Ni:NTA Sepharose (Pierce) gravity flow column with a 2 mL bed volume equilibrated with lysis buffer for immobilized metal affinity chromatography

(IMAC) purification. Bound protein was washed with 50 column volumes (cv) of lysis buffer, followed by 50 cv of a wash buffer containing 300 mM NaCl, 50 mM Tris pH 7.5, 20 mM imidazole, 10% v/v glycerol, 2 mM DTT and 5 mM TCEP. Bound protein was eluted from the column with elution buffer containing 300 mM NaCl, 50 mM Tris pH 7.5, 300 mM imidazole, 10% v/v glycerol, 5 mM DTT and 5 mM TCEP in 2 mL aliquots. Fractions containing protein as determined by Bradford assay [12] were pooled (10 mL total volume) and loaded onto ÄKTA fast protein liquid chromatography instrument (GE Healthcare) with a HiLoad 26/60 Superdex 200 prep grade gel filtration column (GE Healthcare) equilibrated with either 150 mM NaCl, 50 mM Tris pH 7.5, 5% v/v glycerol, 5 mM DTT and 5 mM TCEP for small-angle X-ray scattering (SAXS) analysis or with 150 mM NaCl, 50 mM Tris pH 8.0, 5 mM DTT and 5 mM TCEP for size-exclusion chromatography coupled to a multi-angle light scattering (SEC-MALS) detector. In all cases, SEC purification was performed using a flow rate of 2.5 mL/min. Fractional elutions were assessed with a Coomassie Brilliant Blue R-250 (Bio-Rad) stained 10% Bis-TRIS SDS-PAGE gel with a 1x dilution of MES running buffer (20x stock: 50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3) (Invitrogen) and desired fractions were pooled and stored at 4 °C while concentrating.

2.2.1.2 Concentrating WC I/II Proteins

Collected SEC fractions pooled and stored at 4 °C were loading into a 100

kDa Amicon Ultra-15 Centrifugal Filter Unit (Millipore), pre-equilibrated with 15 to 17 mL of gel filtration buffer (GFB) at 3,000 rpm to remove trace amounts of glycerol applied to the filter membrane during the manufacturing process. Once the membrane was readied, 15 to 17 mL of pooled protein was concentrated at 4 °C and 2,000 rpm. Centrifugation was stopped every 10 min to allow for thorough mixing. This was continued until the volume was reduced to approximately 5 mL and the frequency of mixing was decreased to 5 min. When the volume was reduced to approximately 2.5 to 3 mL, centrifugation was further reduced to 1800 rpm with mixing every 5 min. This continued until protein was brought to the desired concentration as determined by Bradford assay.

As a qualitative measure of the volume reduction, eluent was removed each mixing point. As the desired concentration was approached, a sample of eluent was obtained for later use as a buffer standard in SAXS experiments.

All protein and buffer samples were stored at 4 °C for SAXS experiments or alternatively protein needed for crystallization screening (described in Chapter 1) was aliquoted into 50 µL aliquots and flash frozen in liquid nitrogen and stored at -80 °C.

2.2.2 SEC-MALS Analysis

A mixture of WC I (705-918) and II (174-455) was concentrated to 6.5 mg/mL, as determined by Bradford assay, and subjected to size-exclusion chromatography using a WTC-030S5 column (Wyatt Technology) equilibrated in

150 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM DTT, and 5 mM TCEP. The sizing column was coupled to a Wyatt Technology DAWN HELEOS-II (static 18-angle light scattering detector) and Optilab T-rEX RI (refractive index) detector. The instrument flow rate was set at 1 mL/min while the sample's molar mass and mass distribution was determined with ASTRA software (Wyatt Technology). The samples molecular weight distribution was determined using the Multi-peak Fitting package in Igor Pro (WaveMetrics). Monomeric bovine serum albumin (Pierce) was used as a control and to normalize the light scattering detectors.

2.2.3 SAXS Analysis

2.2.3.1 SAXS Data Collection on F2 Beamline

SAXS analysis of WC proteins was performed at the Cornell High Energy Synchrotron Source (CHESS). Data was collected at the F2 station through the use of a portable, high-throughput sample-changing system that incorporates an automated liquid handling system in conjunction with a quartz capillary flow cell [13]. The F2 station is equipped with multilayer optics that provides a 9.881 keV (1.256 Å) beam collimated to 250 x 250 μM. Sample loading was facilitated by an automated pipetting robot that aspirates protein samples from a 96-well plate chilled to approximately 7°C into a quartz capillary flow cell 10 μm thick and 2 mm in diameter. To minimize radiation damage and subsequent deposition onto the capillary, samples were oscillated in the X-ray beam. Data on the F2 station was collected using a PLATUS 100 K-S detector (Dectris) and calibrated using

the known wavelength and silver behenate powder. Glucose isomerase and lysozyme (Pierce) were used as standards. Sample loading was controlled by in-house software (Robocon).

2.2.3.2 Capillary Cleaning

Prior to obtaining each sample data set, the capillary tube was iteratively washed with three 70 µL water plugs, followed by a 70 µL mild detergent wash solution consisting of 2% m/v solution of N-dodecyl-N,N-(dimethylammino)butyrate, and again three 70 µL water plugs. Once complete the capillary was washed with three 70 µL plug of 150 mM NaCl, 50 mM Tris pH 7.5, 5% v/v glycerol, 5 mM DTT and 5 mM TCEP.

2.2.3.3 Obtaining SAXS Buffer Scattering Profiles

After washing the capillary with buffer, a 30 µL sample of flow-through collected for each protein concentration was used equilibrate the capillary and prevent possible dilution effects. This was in turn followed by a second 30 µL plug of sample for data collection. Each collected flow through was exposed 15 times for 30 s, generating a set of scattering profiles that were checked for radiation damage and comparison to the respective sample data sets. This washing process was repeated for all subsequent buffer and sample exposures.

2.2.3.4 Obtaining SAXS Sample Scattering Profiles

Varying dilutions of WC I and II proteins as described in Chapter 1 were stored for 4 to 7 h at 4 °C prior to centrifugation for 20 min at 4 °C and 14,000

rpm, prior to subsequent beam exposures of 15 times for 30 s. Scattering profiles were regularly checked for protein aggregation, radiation damage and sample deposition on the capillary surface.

2.2.3.5 Analyzing SAXS Data

Collected data was normalized by incident intensity, averaged and plotted using RAW [14]. GNOM [15] was used to calculate the distance distribution function (P(r)), the proteins maximum dimension (Dmax) and the radius of gyration (Rg). Molecular weight estimates were generated through comparison of scattering data to external standards [16], SAXSMOW [17] and estimates of the Porod volume [18].

DAMMIF [19] was used to generate individual molecular envelopes for each sample dilution, which were then averaged using DAMAVER [20] to create consensus envelopes. SITUS [21] (Version 2.7.2) was used to generate bead models from averaged envelopes, which were fitted and visualized with Protein Data Bank (PDB) structures using PYMOL (Molecular Graphics System, Version 1.3 Schrödinger, LLC) and UCSF CHIMERA (Version 1.7) [22] respectively. Fitting of experimental scattering curves to scattering curves generated from known structures was obtained using CRYSOL, Version 2.8 [23].

2.3 Results

2.3.1 Presentation of Sample Data

The discussion of White Collar I and II SAXS experiments begins with *N. crassa* WC I (705-918), (705-1000) and WC II (174-455), (179-500) variants. Analysis for this section provides a more thorough treatment of experimental considerations, with subsequent sections more concisely describing the analysis of copurified PAS domain containing WC I/II (705-918)/(174-455) and PAS with the C-terminal zinc finger domain proteins (705-1000)/(179-500).

2.3.2 WC I and II SAXS Data

Initial efforts to biophysically characterize *N. crassa* WC I and II proteins focused on the individual variants, whose preparation is described in Chapter 1. From these studies, analysis of data collected for WC I (705-1000) and WC II (174-455), (179-500) indicated significant aggregation in all concentrations tested preventing analysis (not shown). Results for WC I (705-918), however, did not show significant aggregation allowing for data processing.

2.3.3 WC I (705-918) SAXS Analysis

2.3.3.1 Appearance of a Sharp Downturn at Small q Values in Data

All scattering profiles obtained, including those for water, standards, buffer and protein samples displayed a sharp downturn at very small q values as seen

in the linear X-linear Y scattering profile for water (Figure 2.3.1). This effect was due to placement of the beamstop at the CHESS F2 beamline and not due to interparticle effects, protein aggregation due to radiation damage or deposition of protein aggregates on the capillary tube during data collection [13]. Evaluation of scattering data indicated the downturn did not have an effect on subsequent data analysis.

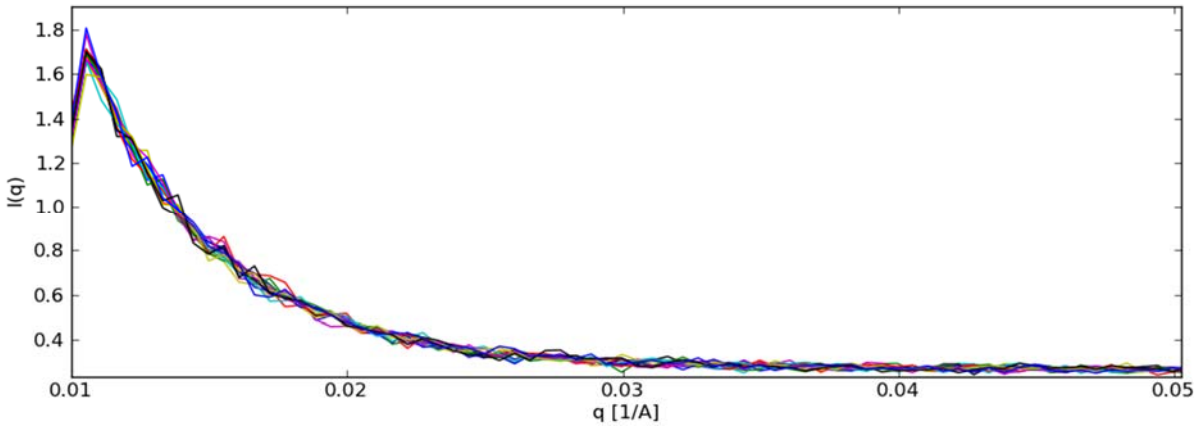

Figure 2.3.1 A Sharp Downturn in Scattering Data is Observed in Small q Values. Experimental setup at the F2 beamline resulted in a sharp downturn in very low q values for all collected scattering profiles as shown in a linear X–linear Y plot of the scattering profile for water.

2.3.3.2 Evaluating WC I (705-918) for Radiation Damage

Buffers and samples were monitored for radiation damage and aggregation from beam exposure by evaluating data using a variety of parameters. The first method was the comparison of 15 sequential 30 s scattering profiles for protein concentrated to 1.2, 2.4 and 4.9 mg/mL. I(q) versus q curves generated using RAW were monitored in linear X–linear Y and log X–

log Y plots [14]. In all cases, scattering curves for each of the successive exposures overlapped with no shifting of q to higher or lower values, indicating the samples were not undergoing significant aggregation nor radiation damage while exposed to the beam [24, 25], as shown in log X–log Y plots (Figure 2.3.1.1), where I(q) is the scattering intensity and $q= 4\pi\sin(\theta/\lambda)$ and is defined as the momentum transfer [26].

2.3.3.3 Generating Interference-free Scattering Curves for WC I (705-918)

To determine if the scattering data obtained for each of the WC I (705-918) samples had negligible interparticle effects and could represent infinitely dilute approximations, buffer data sets were averaged and then subtracted from averaged sample scattering profiles. These curves were then plotted with linear X–log Y axis and scaled using RAW to generate superposition plots with linear X–log Y axis (Figure 2.3.1.2). Comparison of the curves in the superposition plot shows scattering values at very low q values are unchanged, indicating each of the concentrations can be considered infinitely dilute [27]. However, the scattering curves do deviate at higher q values, indicative of a subtle change in the molecules shape, suggestive of a concentration dependent change in the excluded volume, which in turn can affect protein folding [26, 28].

2.3.3.4 Determining WC I (705-918) Rg and I(0) with Guinier Plots

Samples were evaluated for protein aggregation and interparticle

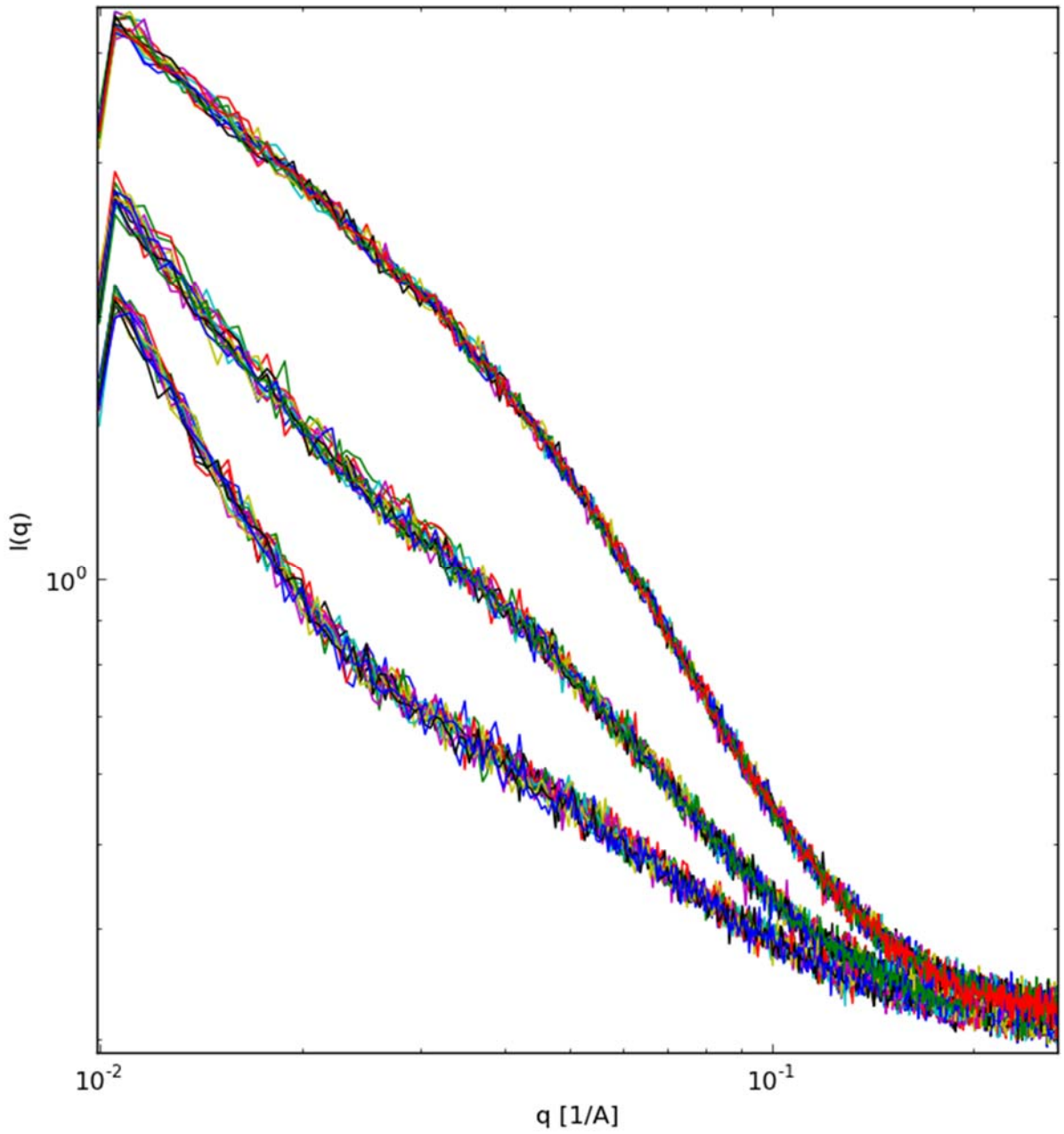

Figure 2.3.1.1 WC I (705-918) Dilutions Show no Change in Scattering Profile from Beam Exposure. Log X–log Y plots of 4.9, 2.4 and 1.2 mg/mL (top, middle, bottom) scattering profiles overlay with no change in q value, indicating protein samples did undergo aggregation or radiation damage from beam exposure.

inference, which appear respectively as up or downturns in intensity at low q values, using Guinier plots [24]. For non-aggregated globular scattering

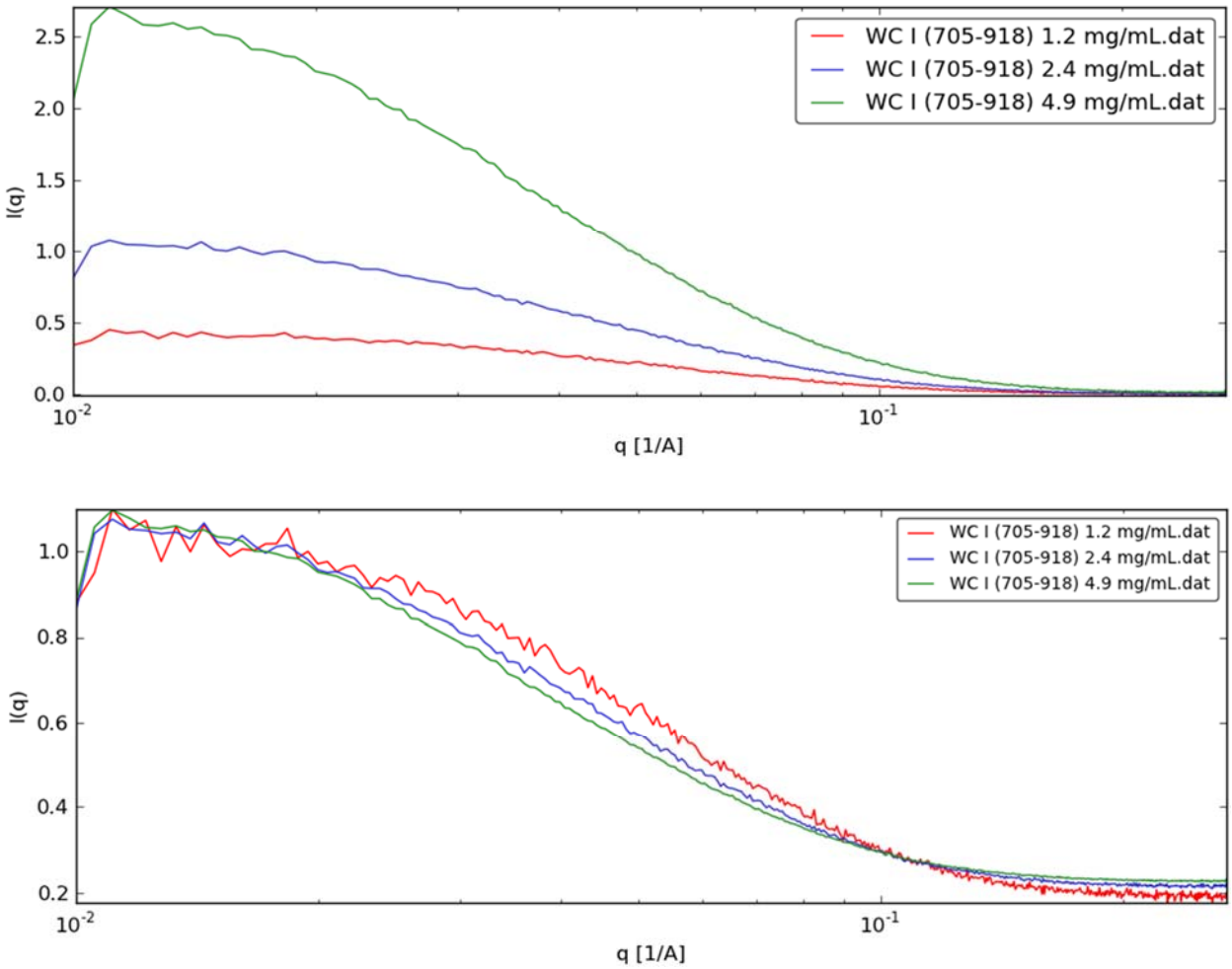

Figure 2.3.1.2 Superpositioned WC I (705-918) Scattering Profiles. Averaged and buffer subtracted scattering profiles (top panel) show relative I(q) values for tested samples. Superpositioned scattering profiles (lower panel) show good alignment at low q values, indicating minimal interparticle interactions in samples concentrated to 1.2, 2.4 and 4.9 mg/mL and suggesting the samples can be considered infinitely dilute.

particles, a linear dependence of log(I(q)) versus q^2 up to a qRg value of 1.3 should be observed [24]. Guinier plots generated with RAW for each WC I (705-918) sample show a linear fit to the scattering data (Figure 2.3.1.3) with calculated values including Rg and I(0) listed in Table 2.3.1.

Table 2.3.1 Values for WC I (705-918) RAW Guinier Plots

Concentration	qStart	qEnd	qRg	Rg	I(0)	rsq
1.2 mg/mL	0.01	0.2689	1.30684	29.81026	0.44273	0.97481
2.4 mg/mL	0.01165	0.03662	1.30605	36.2088	1.12092	0.99229
4.9 mg/mL	0.01221	0.0333	1.30056	39.72326	2.80967	0.99445

2.3.3.5 Evaluating WC I (705-918) Scattering Data with PRIMUS

Additional analysis of scattering data was performed with PRIMUS [29], which uses AUTORG [30] and GNOM [15] to obtain estimates of data quality, pairwise distance distribution function (P(r)), maximum distance (Dmax), Porod volume and Rg and I(0) estimates. Comparison of I(0) and Rg values generated using the P(r) function, show good agreement with values obtained using RAW (Table 2.3.1.1). P(r) functions for WC I (705-918) at 1.2 and 4.9 mg/mL (Figure 2.3.1.4) show a smooth slope near Dmax, however the 2.4 mg/mL sample shows bumps near Dmax, which is indicative of protein aggregation [24, 26].

Additionally, GNOM provides an estimate of data quality and indicates the presence of aggregation, based on the scattering data in the Guinier range. Quality is reported as a percentile ranging from 0 to 100% (unusable to ideal) and aggregation is reported as 1 (with) or 0 (without) [30]. Quality estimates showed a decreasing trend with values at ~80, 70 and 60% for samples concentrated to 1.2, 2.4 and 4.9 mg/mL respectively, with no report of aggregation for any of the tested samples (Table 2.3.1.1).

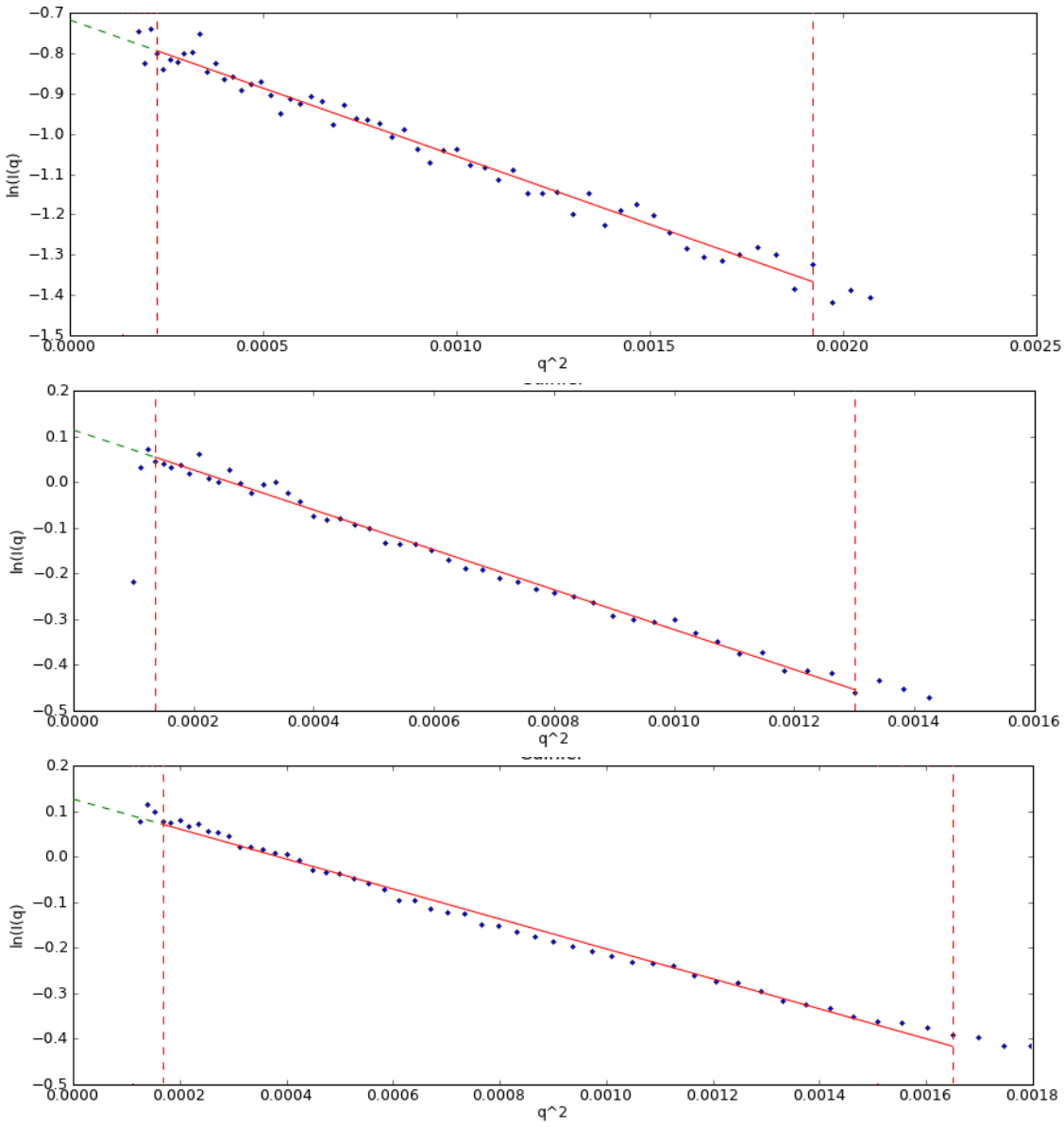

Figure 2.3.1.3 Guinier plots for WC I (705-918). A linear fit for is obtained in the Guinier region for samples at 1.2, 2.4 and 4.9 mg/mL (top, middle and bottom), with no change in intensity at low q values, suggestive of non-aggregated samples.

2.3.3.6 WC I (705-918) Molecular Weight Estimates

Estimates for WC I (705-918) molecular weight were generated in three ways, using concentration dependent and independent methods for comparison

Table 2.3.1.1 AUTORG and GNOM Estimates for WC I (705-918)

	1.2 mg/mL	2.4 mg/mL	4.9 mg/mL
AUTORG			
Rg	29.6 ± 1.353	35.21 ± 1.125	39.411 ± 0.583
I(0)	0.441 ± 0.14	1.098 ± 0.026	2.796 ± 0.026
SRg Limits	0.558 (1.298)	0.684 (1.29)	0.525 (1.29)
Fidelity	1	1	1
Range	16-61	17-48	6-41
GNOM			
Quality Estimate [%]	82.6	70.7	61.5
Reciprocal Space Rg/I(0)	30.31 ± 0.44	36.94 ± 1.11	37.9 ± 2.74
Real Space Rg/I(0)	30.47 ± 0.44	37.23 ± 1.11	39.91 ± 2.74
Porod Volume	62625.8	89909.5	101864
Range	16-467	17-467	6-467
Dmax	101	123.83	130

of predicted values. First, to obtain concentration dependent molecular weight estimates, I(0) values from RAW generated Guinier analysis of WC I protein samples, glucose isomerase (GI) and lysozyme (Lyso) standards (std) of known concentration were compared. In this relationship, $\text{Mol. Wt.}_{(sample)} = ((I(0)_{(sample)}/c_{(sample)}) / (I(0)_{(std)}/c_{(std)}))*\text{Mol. Wt.}_{(std)}$, where I(0) is the forward scattering intensity, c is the concentration in mg/mL, Mol. Wt. is the molecular weight for samples and standards [16].

In addition to molecular weight estimates generated from the comparison of forward scattering data to that of known standards, mass estimates on a relative scale evaluating particle volume were also obtained from analysis of GNOM files using the online SAXSMOW server

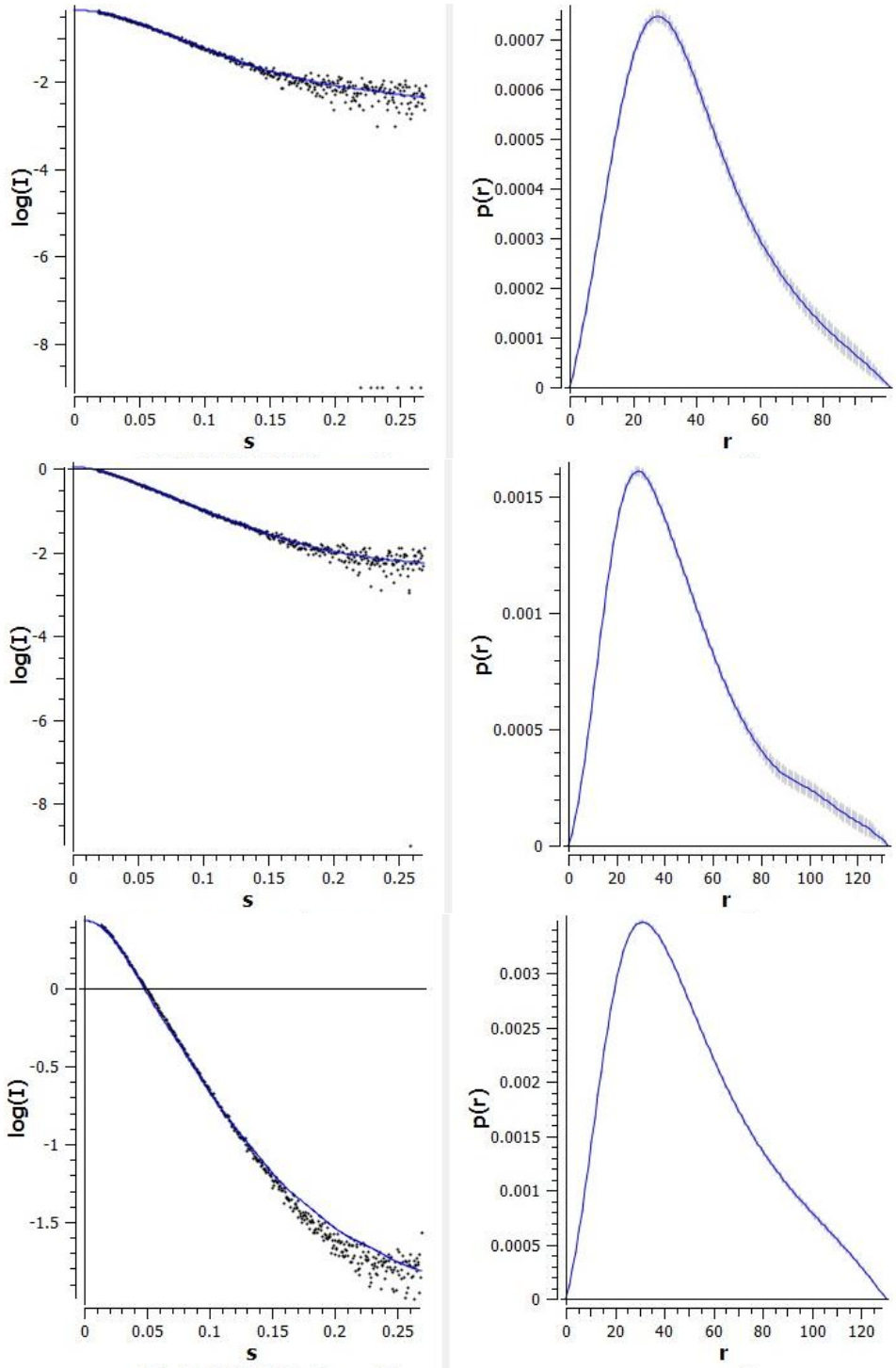

Figure 2.3.1.4 P(r) Plots for WC I (705-918). P(r) functions (right) with accompanying data fits (left) for samples at 1.2 and 4.9 mg/mL (top and bottom) show a smooth slope near Dmax, indicative of non-aggregated samples. Small bumps near Dmax for the 2.4 mg/mL sample (middle) suggest of the onset of aggregation.

(http://www.if.sc.usp.br/~saxs/saxsmow.html). Here empirically determined preset q_{max} ($Å^{-1}$) values of 0.15, 0.20 and 0.25 as provided by the SAXSMOW were selected for approximation of the Porod invariant (Figure 2.3.1.5) [17].

As a final method for comparison, molecular weight for the WC I proteins were estimated from Porod volumes (Vp) analysis as reported by PRIMUS, using an empirical relationship of (Vp*1.2)/2= Mol. Wt. [18]. Estimates using the methods described show a concentration dependent increase in predicted molecular weight (Table 2.3.1.2 and Figure 2.3.1.6).

Depending on a variety of experimental factors, including the accuracy in determining sample and standard concentrations as well as the quality of scattering data, molecular weight estimates obtained from SAXS experiments are reported to have at best an error of approximately 10% [16]. As observed in Table 2.3.1.2, and Figure 2.3.1.6, estimates for the WC I (705-918) samples show significant variation from the predicted molecular weight of a 51 kDa dimer.

Possible explanation for the observed discrepancies in I(0) generated estimates can be attributed to error in the reported sample and standard concentrations, as this method is dependent on accurate assessment of protein concentration [16].

Estimates generated using SAXSMOW and evaluations of the Porod

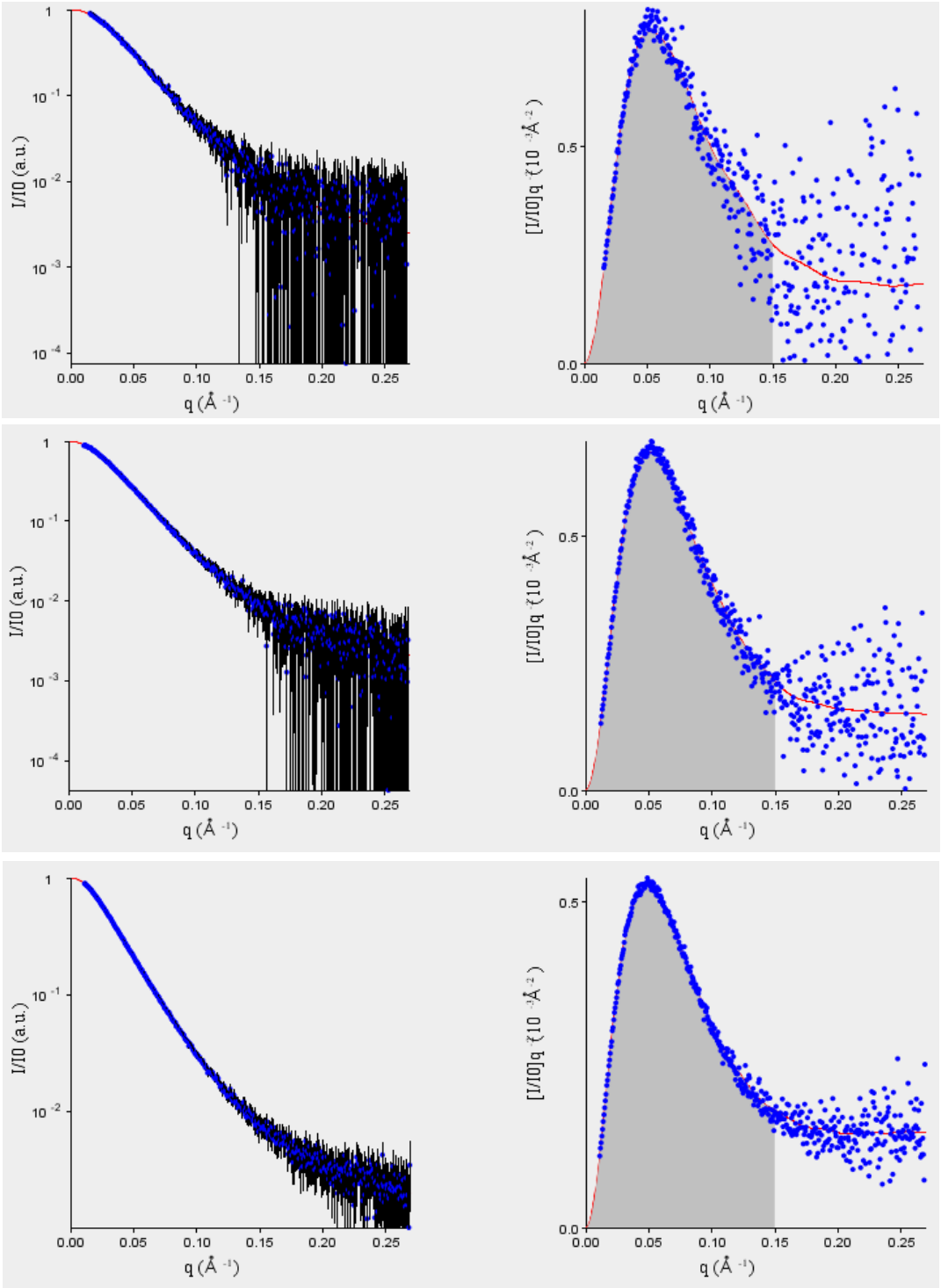

Figure 2.3.1.5 SAXSMOW Analysis of WC I (705-918). Analysis of Porod regions for 1.2, 2.4 and 4.9 mg/mL (top, middle and bottom) using SAXSMOW show a low signal to noise ratio for qmax= 0.2 and 0.25 values, indicating the need for a qmax= 0.15 cutoff value in molecular weight estimations.

Table 2.3.1.2 Molecular Weight Estimates for WC I (705-918)

	I(0) (kDa)		SAXSMOW (kDa)			Vp (kDa)
Concentration	Lyso	GI	0.15q	0.2q	0.25q	(Vp*1.2/2)
1.2 mg/mL	38	41	40	46	53	32
2.4 mg/mL	48	52	63	68	75	45
4.9 mg/mL	59	63	65	71	81	49

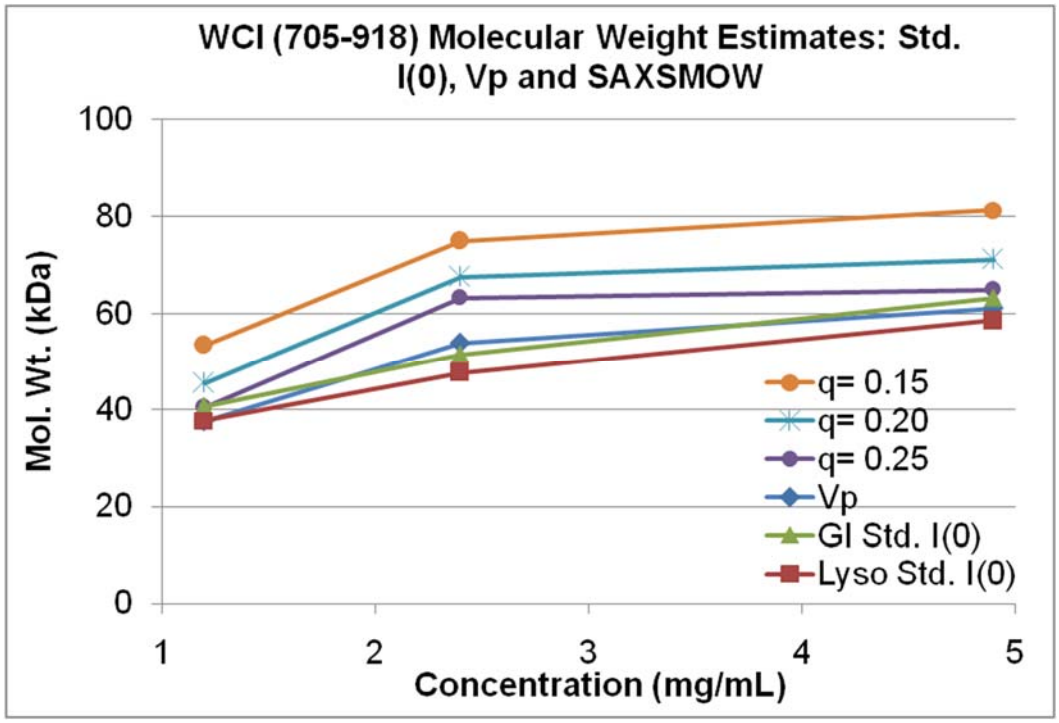

Figure 2.3.1.6 A Concentration Dependent Increase in WC I (705-918) Molecular Weight Estimates is Observed. Comparison of WC I molecular weight estimates from 1) I(0) values to known concentrations of lysozyme (Lyso Std.) and glucose isomerase standards (GI Std.), 2) SAXSMOW estimates based on variation in qmax (qmax= 0.15, 0.20 and 0.25) and 3) analysis of the Porod volume (Vp) as determined by PRIMUS using the relationship (Vp*1.2)/2= Mol. Wt., show a concentration dependent increase in predicted molecular weight.

volume also show deviation from the predicted molecular weight. As seen in

Figure 2.3.1.5, significant error is associated with scattering data at higher q

values for the WC I samples, limiting qmax to 0.15, and in turn making accurate estimation of the predicted Porod volume difficult [17].

2.3.3.7 Concentration Dependent Increases in WC I (705-918) SAXS Data

While significant aggregation was not indicated in superposition, Guinier or P(r) plots for WC I (705-918) scattering data, an increase in I(0), Rg, Dmax, Vp and molecular weight estimates with concentration is observed (Table 2.3.1.1, 2.3.1.2 and Figure 2.3.1.7). These increase are suggestive of either a change in WC I multimerization or concentration dependent aggregation in tested samples [26].

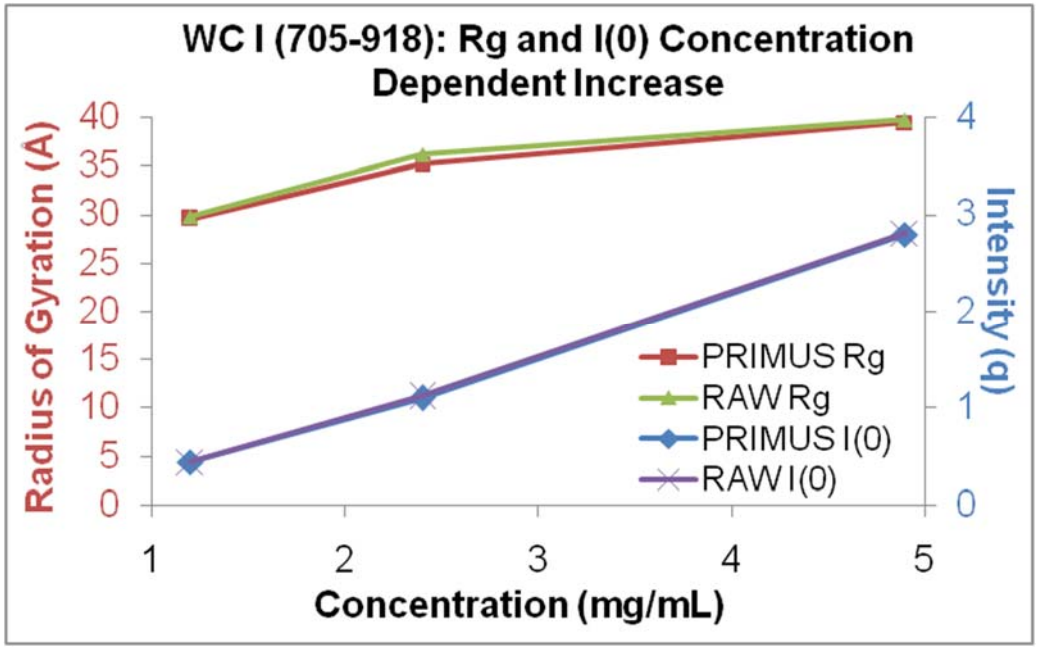

Figure 2.3.1.7 I(0) and Rg Values Show a Concentration Dependent Increases in WC I (705-918) Scattering Data. Comparison of calculated Rg and I(0) values generated using Primus and RAW show an increase for samples at 1.2, 2.4 and 4.9 mg/mL, which is indicative of a change in multimeric state or concentration dependent aggregation.

2.3.3.8 Evaluating WC I (705-918) Folding with Kratky Plots

To evaluate folding and compactness of WC I (705-918), a Kratky plot ($I(q)q^2$ versus q) was generated (Figure 2.3.1.8). From this plot, a broad bell-

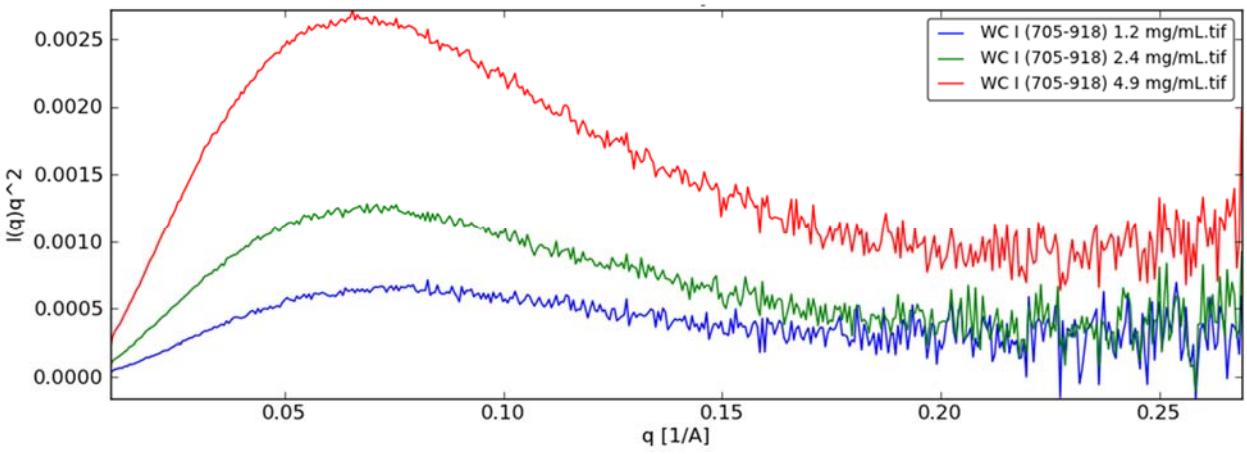

Figure 2.3.1.8 A Kratky Plot Indicates WC I (705-918) is Folded. Evaluation of scattering data shows the presence of a peak at low q and absence of a steady rise in $I(q)q^2$ with increasing q, consistent with a folded protein for all proteins tested.

shaped curve with no increase in $I(q)q^2$ with q is observed for the sample at 4.9 mg/mL, indicative of a folded globular protein with an extended conformation [31]. For samples at 1.2 and 2.4 mg/mL, while a lower in $I(q)q^2$ is observed, neither a plateauing nor an increase in $I(q)q^2$ with q is observed, suggesting that the WC I samples remain folded at lower concentrations.

2.3.3.9 Generating WC I (705-918) *ab initio* Shape Reconstructions

Low resolution three-dimensional *ab initio* shape reconstructions from one-dimensional scattering data were obtained by inputting the P(r) function from

GNOM (output file) into DAMMIF, which incorporates a random number generator and a computationally intensive high temperature annealing method [19]. This process was repeated for each sample 10 times, with and without a P2 symmetry constraint, resulting in shape reconstructions composed of small, tightly packed and interconnected dummy atoms (not shown).

The low resolution *ab initio* models were then submitted to DAMAVER [20] for comparison and subsequent generation of a consensus envelopes. To facilitate this, DAMAVER uses SUPCOMB [32] to iteratively compare and evaluate the models for spatial self-consistency, as defined by a dissimilarity measure termed the normalized spatial discrepancy (NSD). NSD values between 0.4-0.7 are reported as self-consistent and well-behaved, whereas values ≥ 1.3 or those with a NSD greater than 2 standard deviations from the mean are considered poorly behaved and inconsistent.

NSD values for each of the reconstructions were then determined by DAMAVER. To facilitate this, the solutions NSD values were calculated. The lowest determined NSD was then selected as a reference model and based on this starting point, an averaged envelope was generated. As seen in Table 2.3.1.3, all of the DAMMIF solutions, were considered stable and used to generate an averaged envelope.

Average NSD values for consensus envelopes and their variations (ΔNSD) indicate solutions for each of the tested WC I (705-918) samples are well

Table 2.3.1.3 Statistics for WC I (705-918) *ab initio* DAMMIF Shape Reconstructions

WC I (705-918): DAMMIF Average NSD						
	1.2 mg/mL		2.4 mg/mL		4.9 mg/mL	
PDB #	P1	P2	P1	P2	P1	P2
1	0.6	0.539	0.765	0.619	0.758	0.755
2	0.61	0.542	0.765	0.625	0.777	0.77
3	0.618	0.551	0.772	0.639	0.793	0.777
4	0.619	0.558	0.776	0.651	0.809	0.787
5	0.621	0.591	0.778	0.665	0.809	0.795
6	0.625	0.591	0.785	0.673	0.814	0.798
7	0.629	0.616	0.804	0.703	0.854	0.805
8	0.636	0.674	0.807	0.718	0.856	0.823
9	0.637	0.728	0.808	0.723	0.86	0.851
10	0.651	0.733	0.815	0.779	0.881	0.854

behaved and self-consistent (Table 3.3.1.4), with averaged envelopes generated from overlapping dummy atoms of a constant sphere size as visualized using PYMOL (Figure 2.3.1.9).

From these models, basic features in the envelopes with and without the P2 symmetry constraint are conserved. Additionally, a subtle increase in envelope dimensions was observed with increasing concentration.

Table 2.3.1.4 NSD Values for WC I (705-918) Consensus Envelopes

WC I (705-918): Average NSD						
	1.2 mg/mL		2.4 mg/mL		4.9 mg/mL	
	P1	P2	P1	P2	P1	P2
Average NSD	0.625	0.612	0.788	0.68	0.821	0.802
ΔNSD	0.015	0.074	0.019	0.051	0.04	0.033

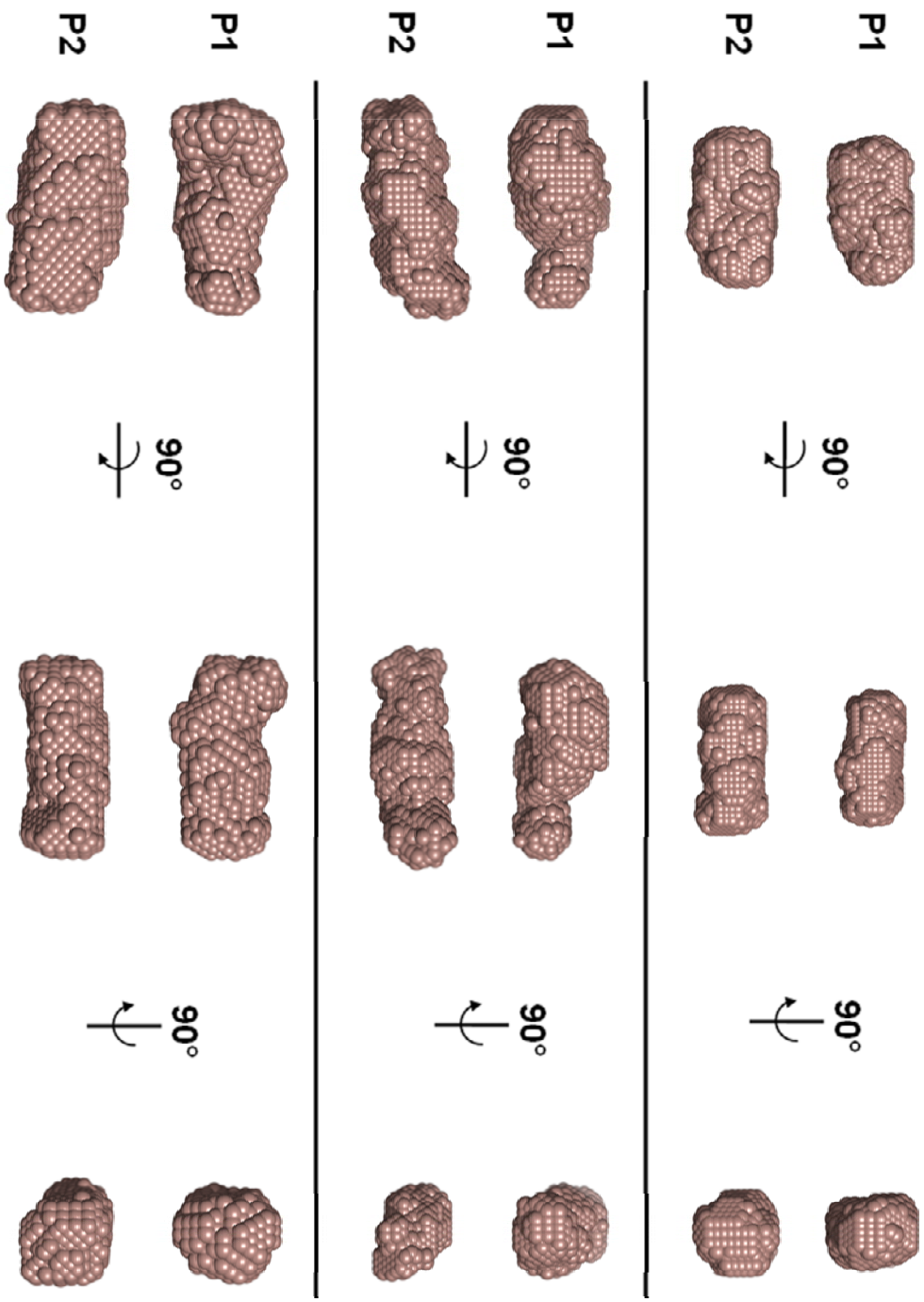

Figure 2.3.1.9 Averaged Low Resolution *ab initio* Shape Reconstructions for WC I (705-918). Averaged consensus envelopes for 1.2, 2.4, 4.9 mg/mL (top, middle, bottom) show a slight increase in dimensions with increased concentration. Also observed is a basic retention of symmetrical features for models with and without a P2 symmetry constraint.

2.3.4 Copurifying *N. crassa* WC I/II (705-918/174-455)

Previously purified *N. crassa* WC I (705-918) and WC II (174-455) proteins were mixed at 0.75 and 1.5 mg/mL with equimolar concentrations and stoichometric ratios, for both proteins, ranging from 4:1 to 1:1 for crystallization screening. In all cases protein solutions remained clear with no visible change upon mixing. However, when WC I (705-1000) was mixed with WC II (179-500), the solution visibly precipitated at all concentrations tested, even when the buffer composition was identical. This observation in concert with the uncertainty associated with the required ratios and concentrations needed to facilitate a WC I/II PAS–PAS domain interaction drove the need for another purification method.

To address this, copurification of WC PAS domain containing proteins, I (705-918) and II (174-455), without the C-terminal zinc finger domains were first undertaken as they consistently demonstrated higher soluble protein expression levels and increased solution stability. Once optimized, the method was applied and adapted as necessary for WC PAS domain containing proteins with the C-terminal zinc finger domains, I (705-1000) and II (179-500).

2.3.4.1 WC I/II (705-918/174-455) Copurification

WC I (705-918) and WC II (174-455) proteins were expressed, harvested and stored separately. Once obtained, cell pellets of approximately 12.5 g were resuspended in lysis buffer, mixed and the overexpressed proteins were purified

together, first with IMAC followed by SEC. The principle elution peak from the SEC column was at 181 mL (Figure 2.3.2), and when compared to standards suggested a species with an apparent molecular weight of approximately 130 kDa, larger than the predicted theoretical molecular weight of 119 kDa for WC I/II complex. This elution volume of 181 mL differed substantially from previous purification of the individual WC I (705-918) and II (174-455) proteins, which eluted at 210 and 199 mL, respectively as shown in Chapter 1 (Figures 1.3.2.2 and 1.3.2). Analysis of SEC elution fractions with a Coomassie stained 4-12% SDS-PAGE gel (Figure 2.3.2) shows bands with molecular weights consistent with the overexpressed WC I/II proteins.

Quantification of relative band intensity indicated an excess of WC II for the early portion of the primary elution peak transitioning to an excess of WC I in the later elution volume. Densitometric analysis of the SDS-PAGE gels loaded with SEC fractions selected for concentration by Syngene GeneTools show an approximate excess of 20% WC II in lane 1 transitioning to an equal ratio in lane 4 for collected fractions (Table 2.3.1.5).

Table 2.3.1.5 Densitometric Analysis of WC I/II (705-918)/(174-455) SEC Elution Fractions Selected for Concentration

Protein	Lane 1	Lane 2	Lane 3	Lane 4
WC II	56%	54%	54%	50%
WC I	44%	46%	46%	50%

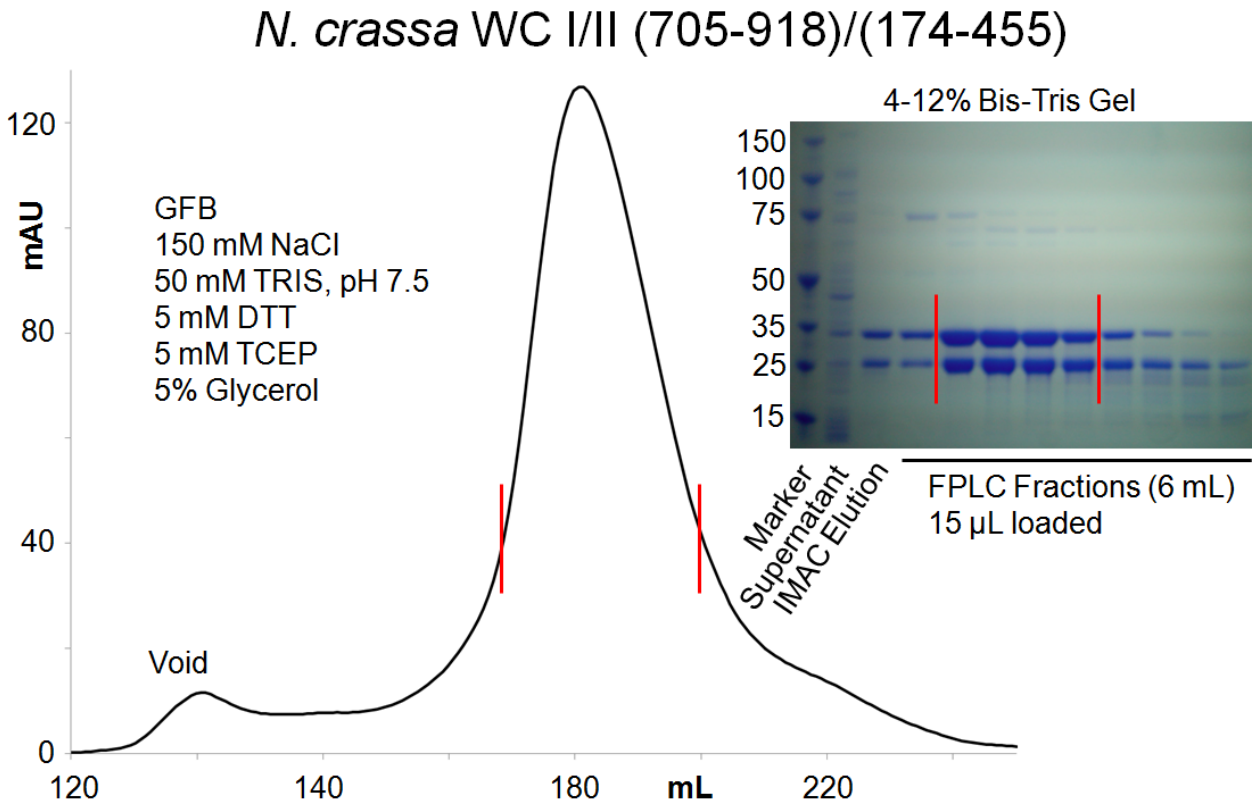

Figure 2.3.2 Copurification of WC I/II (705-918)/(174-455). SEC elution profile of copurified WC I/II (705-918)/(174-455) indicates a species with a hydrodynamic radius consistent with the coelution of WC I/II proteins. SDS-PAGE gel analysis of collected SEC elution fractions shows the presence of both proteins in the primary elution fractions. Red bars on elution the profile correspond to the fractions selected for concentration and SAXS analysis.

A consistent feature in the purification of WC I/II (705-918)/(174-455) proteins is formation of high molecular weight bands appearing between 60 to 75 kDa on SDS-PAGE gels even after heating samples to 95 °C for 5 min in the presence of 10 mM DTT. These bands are attributed to formation of irreversible protein aggregates and were minimized by maintaining a temperature of 4 °C during handling, frequent mixing during shortened sonication rounds, increased

reductant concentrations and performing SEC immediately after IMAC purification.

2.3.4.2 Concentration of WC I/II (705-918)/(174-455)

In previous concentration efforts, WC I and II SEC elution fractions readily produced white-precipitate in the solution with subsequent SDS-PAGE analysis of these samples showing high molecular weight bands. To minimize formation of irreversible aggregates during concentration, SEC fractions in GFB were pooled and concentrated using a stepwise reduction in the centrifugation rate coupled with frequent mixing. Protein samples were concentrated to 0.8, 1.7 and 3.3 mg/mL, as determined by Bradford assay, for subsequent SAXS analysis.

2.3.4.3 Determination of WC I/II (705-918)/(174-455) Storage Stability

To assess the stability of WC I/II (705-918)/(174-455) in GFB, freeze-thaw and storage stability studies were conducted. For the freeze-thaw study, WC I/II (705-918)/(174-455) proteins concentrated to 6.5 mg/mL were aliquoted and flash-frozen in liquid nitrogen, then thawed on ice and evaluated using SDS-PAGE analysis. As shown in Figure 2.3.2.1, a single freeze-thaw cycle was sufficient to induce aggregation, indicating the protein should not be frozen prior to SAXS experiments as the scattering intensity, I(0), is proportional to square of the molecular weight and thus dominates the scattering signal [26].

Previous SDS-PAGE gel analysis of purified WC I/II stored overnight in a

Figure 2.3.2.1 Flash-freezing Induces Aggregation in WC I/II (705-918)/(174-455). SDS-PAGE gel analysis of WC I/II (705-918)/(174-455) concentrated to 6.5 mg/mL, before (lane 1) and after (lane 2) a freeze-thaw cycle indicate one round of freezing is sufficient to induce high molecular weight aggregation.

variety of buffer conditions at 4 °C showed aggregate formation. To evaluate the effect of short-term storage at 4 °C, samples concentrated to 6.5 mg/mL in GFB were taken every 4 h and monitored for aggregate formation with SDS-PAGE analysis. Results indicated high molecular weight band formation was negligible for a period up to 12 h (not shown). Thus, unfrozen protein was stable at 4 °C on the time scale needed for subsequent SAXS analysis.

2.3.5 WC I/II (705-918)/(174-455) SAXS Analysis

2.3.5.1 Evaluating WC I/II (705-918)/(174-455) for Radiation Damage

To evaluate protein stability, 15 sequential 30 s scattering profiles for protein concentrated to 0.8, 1.7 and 3.3 mg/mL were monitored during data collection. From RAW generated log X–log Y plots no shifting of q to higher or

lower values in (Figure 2.3.2.2), indicating the samples were not undergoing aggregation or radiation damage while exposed to the beam [24].

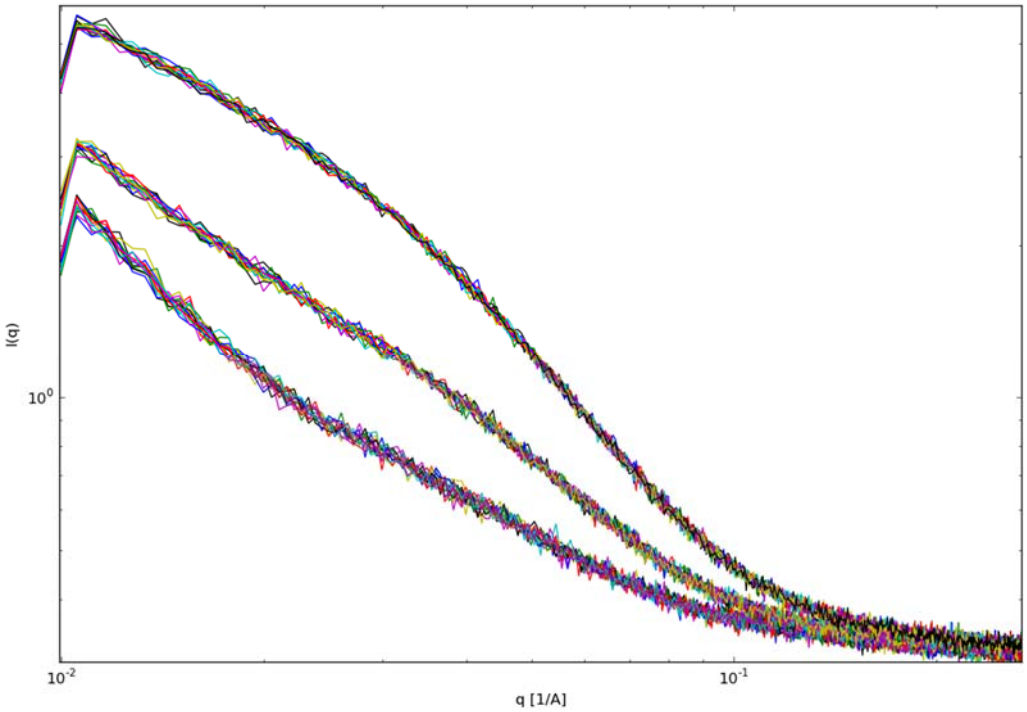

Figure 2.3.2.2 Scattering Profiles for WC I/II (705-918)/(174-455) Dilutions Show No Shifting from Beam Exposure. Log X–log Y plots for WC I/II (705-918)/(174-455) at 3.3, 1.7 and 0.8 mg/mL (top, middle and bottom) overlay with no change in q value, indicating protein samples did undergo aggregation or radiation damage from beam exposure.

2.3.5.2 Generating Interference-free Scattering Curves for WC I/II (705-918)/(174-455)

To verify that SAXS data collected for each of the WC I/II (705-918)/(174-455) samples had negligible interparticle effects and represented infinitely dilute approximations, buffer subtracted, averaged scattering profiles were first generated and scaled using RAW, then superposed with linear X–log Y axis

(Figure 2.3.2.3). From this comparison the scattering values at low q values for each of the concentrations were unchanged, indicating the samples can be considered infinitely dilute [27]. Also observed is deviation of the 3.3 mg/mL

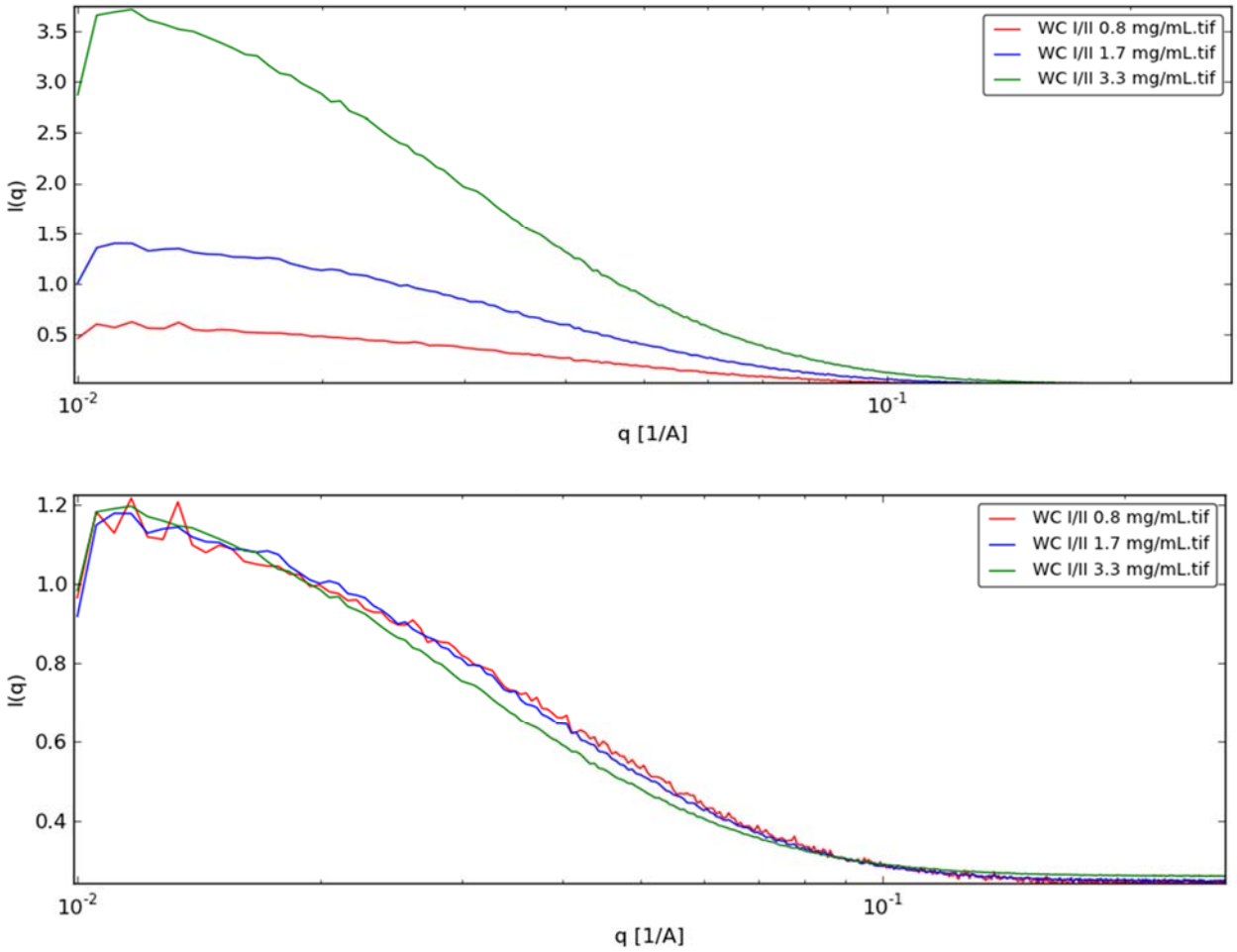

Figure 2.3.2.3 Superpositioned WC I/II (705-918)/(174-455) Scattering Profiles Show No Aggregation. Averaged and buffer subtracted WC I/II (705-918)/(174-455) scattering profiles show relative I(q) (top panel) and superpositioned profiles (lower panel) do not indicate the occurrence of interparticle interactions and samples at 0.8, 1.7 and 3.3 mg/mL can be considered as infinitely dilute.

scattering curve at higher q values, which suggests that the shape of the molecule is subtly changed with increasing concentration [26].

2.3.5.3 Determining WC I/II (705-918)/(174-455) Rg and I(0) with Guinier Plots

Guinier analysis of the WC I/II (705-918)/(174-455) samples for aggregation or interparticle showed no up or downturns in intensity at low q values. Additionally, a linear dependence of scattering data up to a qRg value of 1.3 is observed (Figure 2.3.2.4) with calculated values including Rg and I(0) listed in Table 2.3.1.6.

Table 2.3.1.6 Values for WC I/II (705-918)/(174-455) RAW Guinier Plots

Concentration	qStart	qEnd	qRg	Rg	I(0)	rsq
0.8 mg/mL	0.01387	0.03274	1.29856	40.34573	0.60297	0.9885
1.7 mg/mL	0.01221	0.03052	1.30312	43.4864	1.48107	0.99459
3.3 mg/mL	0.01221	0.02608	1.29733	51.03138	4.09258	0.99685

2.3.5.4 Evaluating WC I/II (705-918)/(174-455)Scattering Data with PRIMUS

Additional analysis of scattering data with AUTORG [30] and GNOM [15], which are part of the PRIMUS [29] suite, showed good agreement with values obtained using RAW (Table 2.3.1.7). Quality estimates for each of the samples were all reported above 70%, which is noted as good quality estimates.

P(r) functions for each WC I/II (705-918)/(174-455) sample showed a smooth slope near Dmax, lacking bumps which are indicative of protein aggregation [24, 26] (Figure 2.3.2.5).

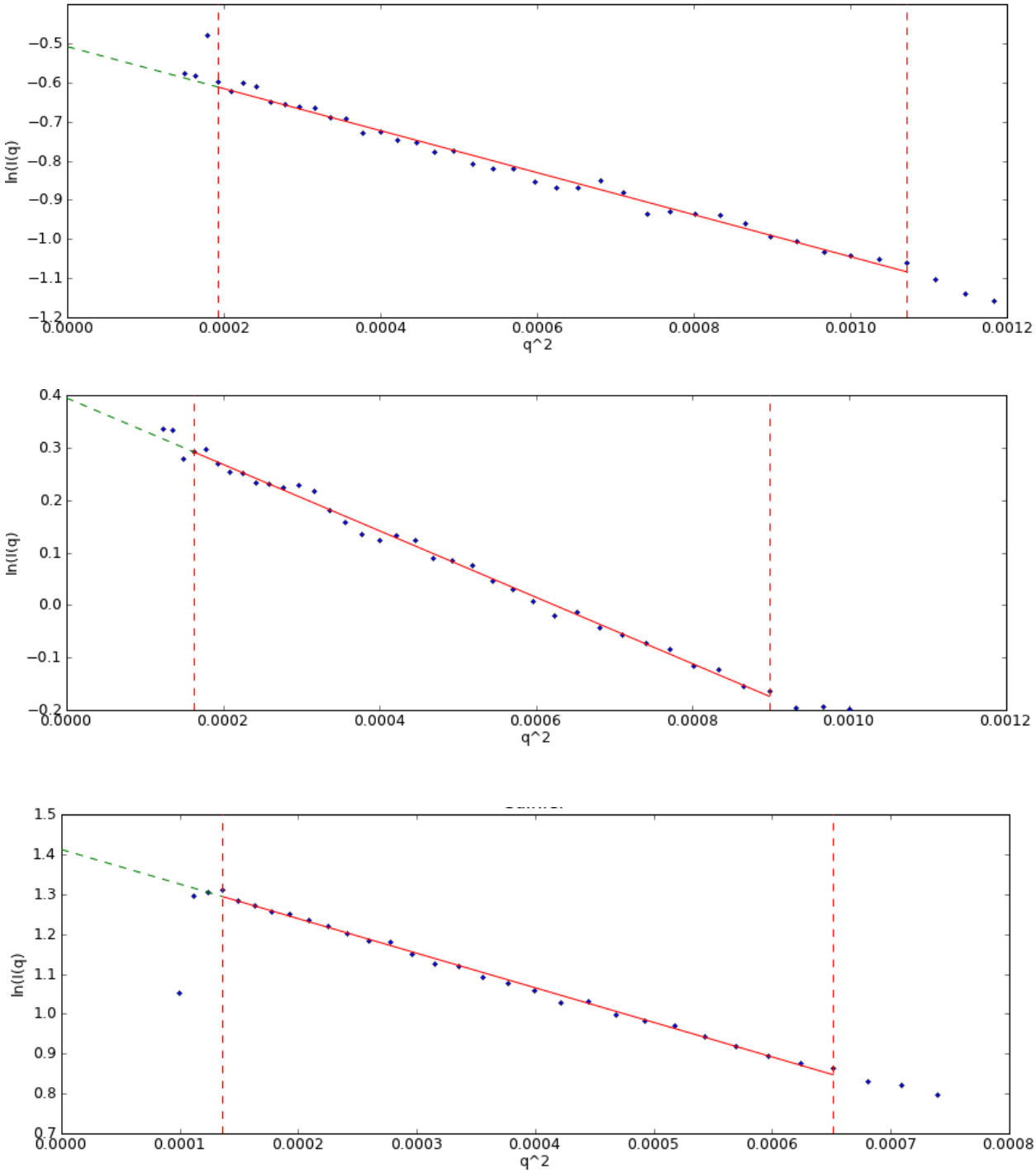

Figure 2.3.2.4 Guinier plots for WC I/II (705-918)/(174-455) Show A Linear Fit. A linear fit for is obtained in the Guinier region for WC I/II (705-918)/(174-455) samples at 0.8, 1.7 and 3.3 mg/mL (top, middle and bottom), with no change in intensity at low q values.

Table 2.3.1.7 AUTORG and GNOM Estimates for WC I/II (705-918)/(174-455)

	0.8 mg/mL	1.7 mg/mL	3.3 mg/mL
AUTORG			
Rg	39.599 ± 6.309	43.672 ± 0.804	53.148 ± 5.469
I(0)	0.594 ± 0.004	1.485 ± 0.007	4.182 ± 0.017
SRg Limits	0.637 (1.275)	0.557 (1.284)	0.619 (1.091)
Fidelity	0.743	0.924	0.831
Range	11-40	5-35	3-19
GNOM			
Quality Estimate [%]	83.1	73.7	72.9
Reciprocal Space Rg/I(0)	40.81 ± 0.6	45.68 ± 1.5	53.1 ± 4.14
Real Space Rg/I(0)	40.99 ± 0.6	45.94 ± 1.5	53.49 ± 4.14
Porod Volume	158370	198049	256222
Range	11-467	5-467	3-467
Dmax	135	165	189

2.3.5.5 WC I/II (705-918)/(174-455) Molecular Weight Estimates

Molecular weight estimates for WC I/II (705-918)/(174-455) were generated from I(0) values derived from RAW analysis of the Guinier region [16], the SAXSMOW online web server, which estimates mass by evaluating particle volumes [17] and mass estimates from evalution of the caclulated Porod volumes using a relationship of (Vp*1.2)/2= molecular weight [18] (Figure 2.3.2.6).

As seen in Table 2.3.1.8 and Figure 2.3.2.7, estimates using the methods described showed a concentration dependent increase in predicted molecular weight for each sample, with significant deviation from the predicted molecular weight of 119 kDa.

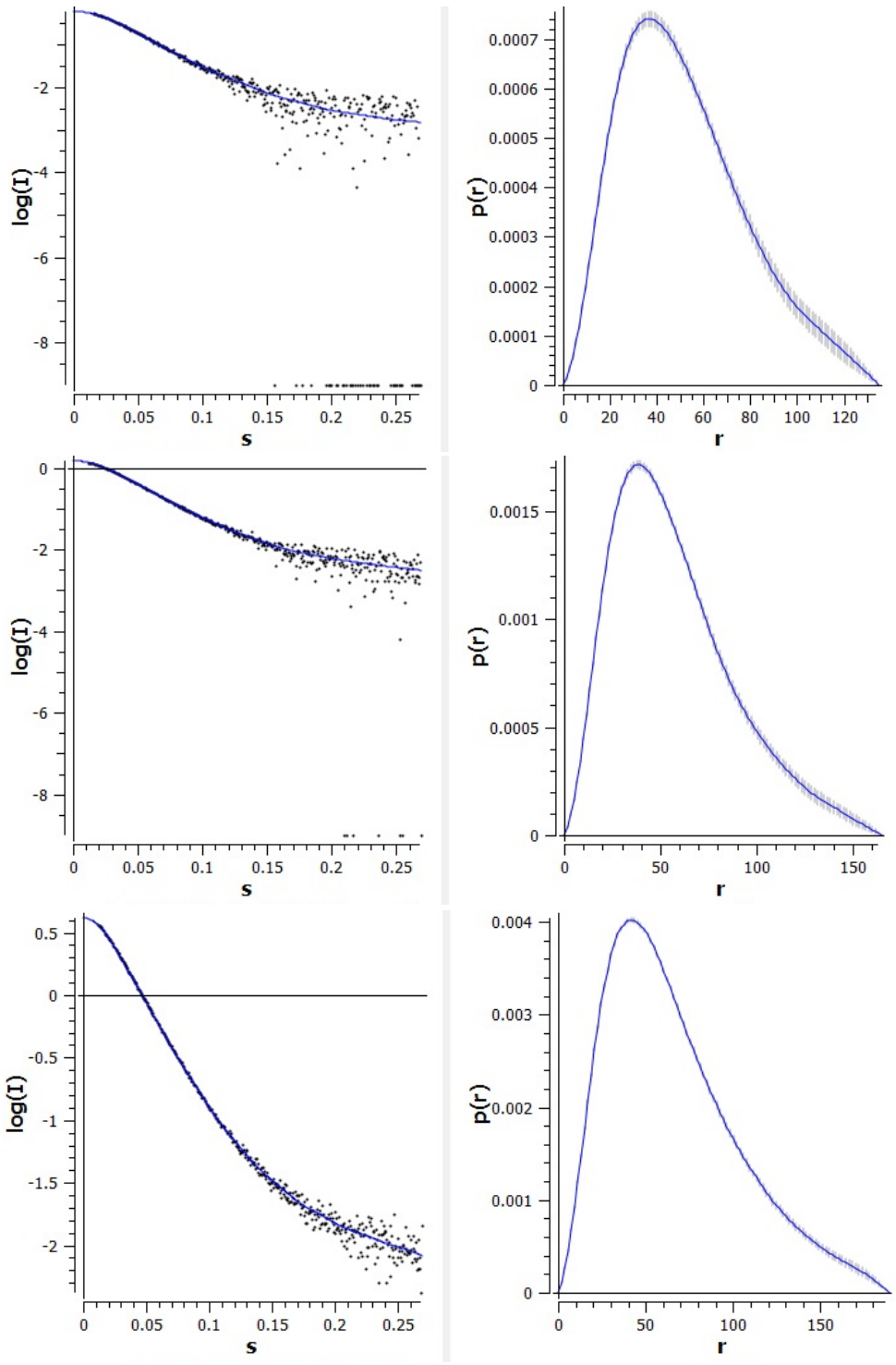

Figure 2.3.2.5 P(r) Plots for WC I/II (705-918)/(174-455). P(r) functions (with fit on left) for samples at 0.8, 1.7 and 3.3 mg/mL (top, middle and bottom) show a smooth slope near Dmax, indicative of non-aggregated samples.

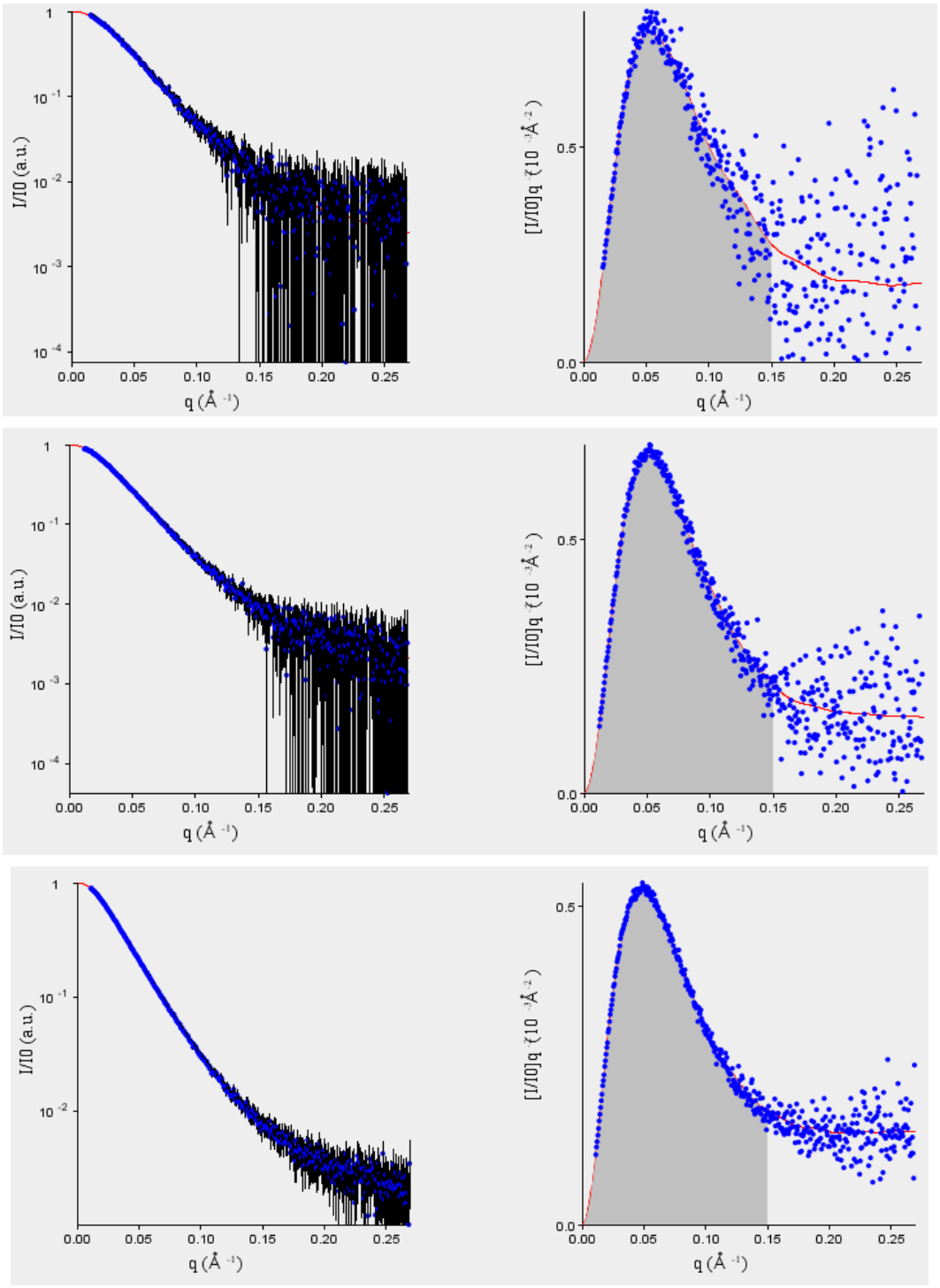

Figure 2.3.2.6 SAXSMOW Analysis of WC I/II (705-918)/(174-455). SAXSMOW analysis of Porod regions for 0.8, 1.7 and 3.3 mg/mL (top, middle and bottom) show a low signal to noise ratio, indicating the need for a qmax= 0.15 cutoff value in molecular weight estimates.

Table 2.3.1.8 Molecular Weight Estimates for WC I/II (705-918)/(174-455)

	I(0) (kDa)		SAXSMOW (kDa)			Vp (kDa)
Concentration	Lysozyme	Glucose Isomerase	0.25q	0.2q	0.15q	(Vp*1.2/2)
0.7 mg/mL	76	82	169	184	203	173
1.2 mg/mL	120	128	193	210	234	195
2.3 mg/mL	132	141	215	235	261	223

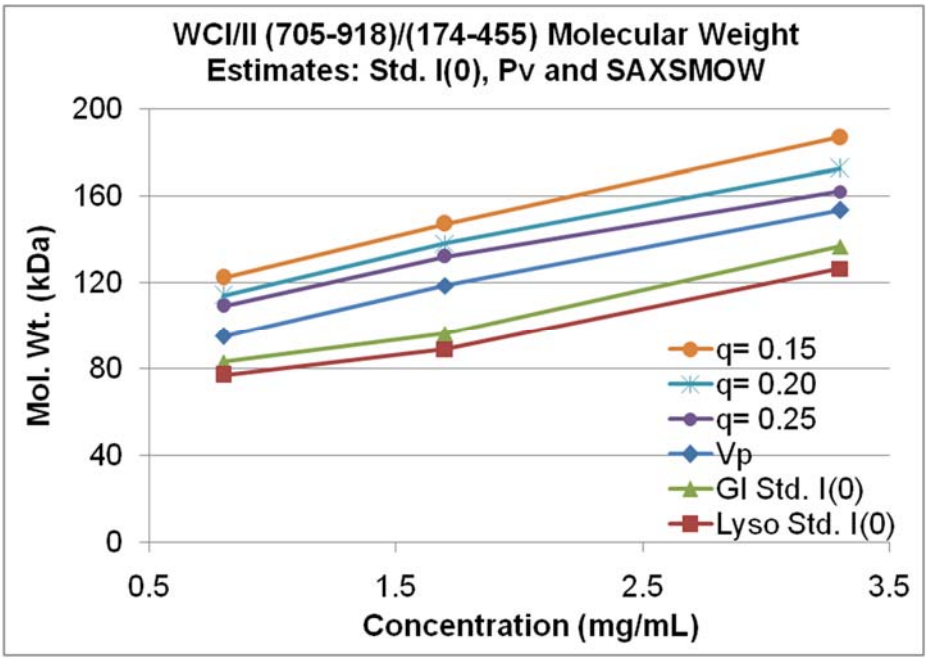

Figure 2.3.2.7 Molecular Weight Estimates for WC I/II (705-918)/(174-455) Show a Concentration Dependent Increase. Comparison of WC I/II molecular weight estimates from 1) I(0) values to known concentrations of lysozyme (Lyso Std.) and glucose isomerase standards (GI Std.), 2) SAXSMOW estimates based on variation in qmax (qmax= 0.15, 0.20 and 0.25) and 3) analysis of the Porod volume (Vp) as determined by PRIMUS using the relationship (Vp*1.2)/2= Mol. Wt., show a concentration dependent increase in predicted molecular weight.

2.3.5.6 Concentration Dependent Increases in WC I/II (705-918)/(174-455) SAXS Data

As seen in superposition, Guinier and P(r) plots of WC I/II (705-918)/(174-

455) scattering data, significant aggregation was not indicated. However, an increase in I(0), Rg, Dmax, Vp and molecular weight estimates with concentration was observed (Table 2.3.1.7, 2.3.1.8 and Figure 2.3.2.8), which is suggestive of either a change in WC I/II multimerization or concentration dependent aggregation in tested samples [26].

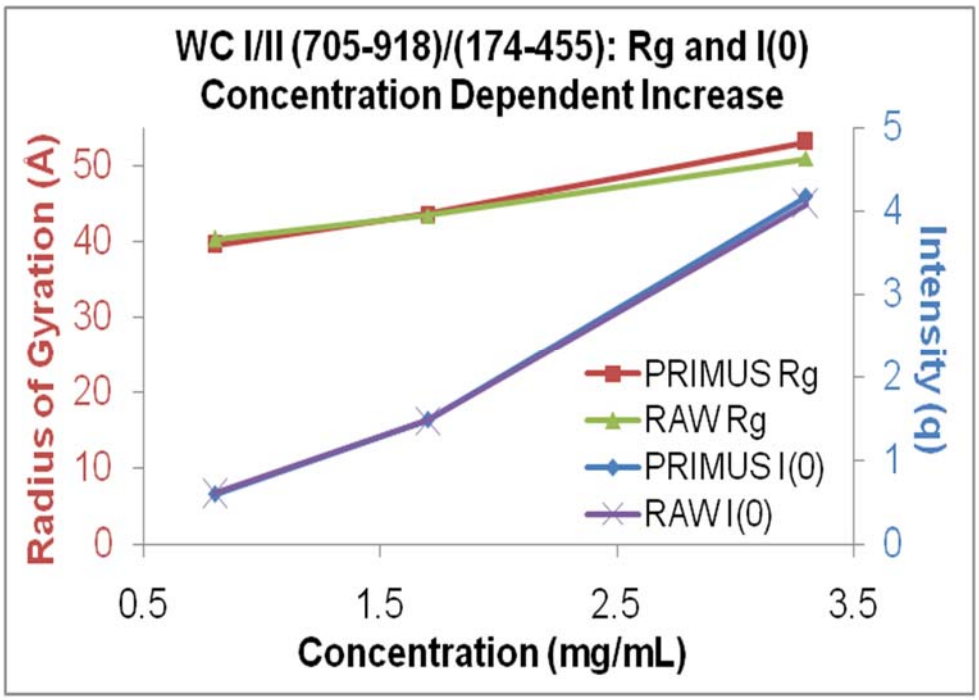

Figure 2.3.2.8 I(0) and Rg Values Show a Concentration Dependent Increases in WC I/II (705-918)/(174-455) Samples. Comparison of calculated Rg and I(0) values from Primus and RAW show a concentration dependent increase for samples at 0.8, 1.7 and 3.3 mg/mL, which is indicative of a change in multimeric state or concentration dependent aggregation.

2.3.5.7 Evaluating WC I/II (705-918)/(174-455) Folding with Kratky Plots

To evaluate folding and compactness, a Kratky plot of WC I/II (705-918)/(174-455) was generated from scattering data (Figure 2.3.2.9). From this

plot, each concentration tested showed a parabolic curve followed by constant $I(q)q^2$ with q, indicative of a folded globular protein an extended conformation [31]. Also noted is a decrease in $I(q)q^2$ at lower concentrations, however, neither a plateauing nor an increase in $I(q)q^2$ with q is observed, further suggesting that the WC I/II samples are folded.

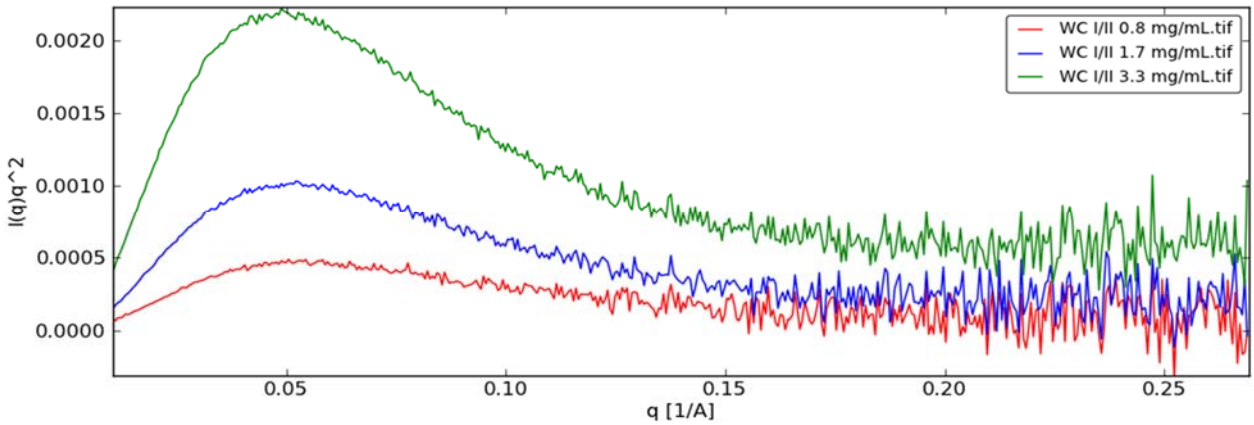

Figure 2.3.2.9 A Kratky Plot for WC I/II (705-918)/(174-455) Shows Folded Protein. Evaluation of scattering data shows the presence of a peak at low q and absence of a steady rise in $I(q)q^2$ with increasing q, consistent with a folded protein.

2.3.5.8 Generating WC I/II (705-918)/(174-455) *ab initio* Shape Reconstructions

Low resolution three-dimensional *ab initio* shape reconstructions of the WC I/II (705-918)/(174-455) samples, with and without a symmetry constraint, were obtained using DAMMIF and DAMAVER. For DAMMIF processing, the P(r) function from GNOM was used to generate 10 low resolution shape constructions with a P1 and P2 symmetry constraint (individual models not shown) [19]. These models were then submitted to DAMAVER [20] and iteratively compared for

spatial self-consistency, resulting in subsequent generation of consensus envelopes.

Results for the analysis of the models spatial self-consistency, as defined by a dissimilarity measure, termed the normalized special discrepancy (NSD), for individual envelopes and consensus envelopes is shown in Table 2.3.1.9 and 2.3.2.

From this analysis, all but two DAMMIF solutions for the 0.8 mg/mL samples, as denoted by bold values with an asterisk, were considered stable and used to generate an averaged envelope Table 2.3.1.9. Reported average NSD values for consensus envelopes and their variations (ΔNSD) indicate solutions for each of the tested WC I/II (705-918)/(174-455) samples are well behaved and self-consistent (Table 2.3.2).

Table 2.3.1.9 Statistics for WC I/II (705-918)/(174-455) *ab initio* DAMMIF Shape Reconstructions

	WC I/II (705-918)/(174-455) Average NSD					
	0.8 mg/mL		**1.7 mg/mL**		**3.3 mg/mL**	
PDB #	P1	P2	P1	P2	P1	P2
1	0.596	0.626	0.67	0.709	0.58	0.546
2	0.619	0.614	0.693	0.712	0.581	0.553
3	0.629	0.623	0.693	0.713	0.583	0.556
4	0.63	0.635	0.699	0.723	0.591	0.562
5	0.636	0.662	0.713	0.73	0.596	0.602
6	0.642	0.664	0.718	0.735	0.601	0.623
7	0.674	0.68	0.72	0.735	0.611	0.65
8	0.682	0.686	0.721	0.744	0.639	0.653
9	0.697	0.704	0.747	0.786	0.649	0.671
10	**1.086***	**0.867***	0.784	0.842	0.651	0.68

Table 2.3.2 NSD Values for WC I/II (705-918)/(174-455) Consensus Envelopes

	WC I/II (705-918)/(174-455) Average NSD					
	0.8 mg/mL		1.7 mg/mL		3.3 mg/mL	
	P1	P2	P1	P2	P1	P2
Average	0.689	0.676	0.716	0.743	0.608	0.61
ΔNSD	0.143	0.074	0.032	0.041	0.028	0.053

Averaged envelopes with overlapping dummy atoms of a constant sphere size were visualized using PYMOL (Figure 2.3.3). From these models, basic features in the envelopes with and without the P2 symmetry constraint are conserved. Additionally, a subtle increase in envelope dimensions is observed with increasing concentration.

2.3.5.9 Superposition of a Canonical PAS Structure with the WC I/II (705-918)/(174-455) Consensus Envelope

Averaged DAMAVER envelopes were further evaluated through superposition with a known canonical PAS fold. To facilitate this, an internal fitting algorithm in UCSF CHIMERA [22] was used to create a superposition of the Drosophila Period structure (3RTY) with the averaged WC I/II (705-918)/(174-455) 0.8 mg/mL envelope.

The Period structure (229-575) is formed by two PAS AB dimers displaying a conical PAS–PAS repeat motif [33] that has sequence homology to the two WC I/II PAS domain containing proteins. As shown in Figure 2.3.3.1, good agreement is observed between the mesh representation of the WC I/II (705-

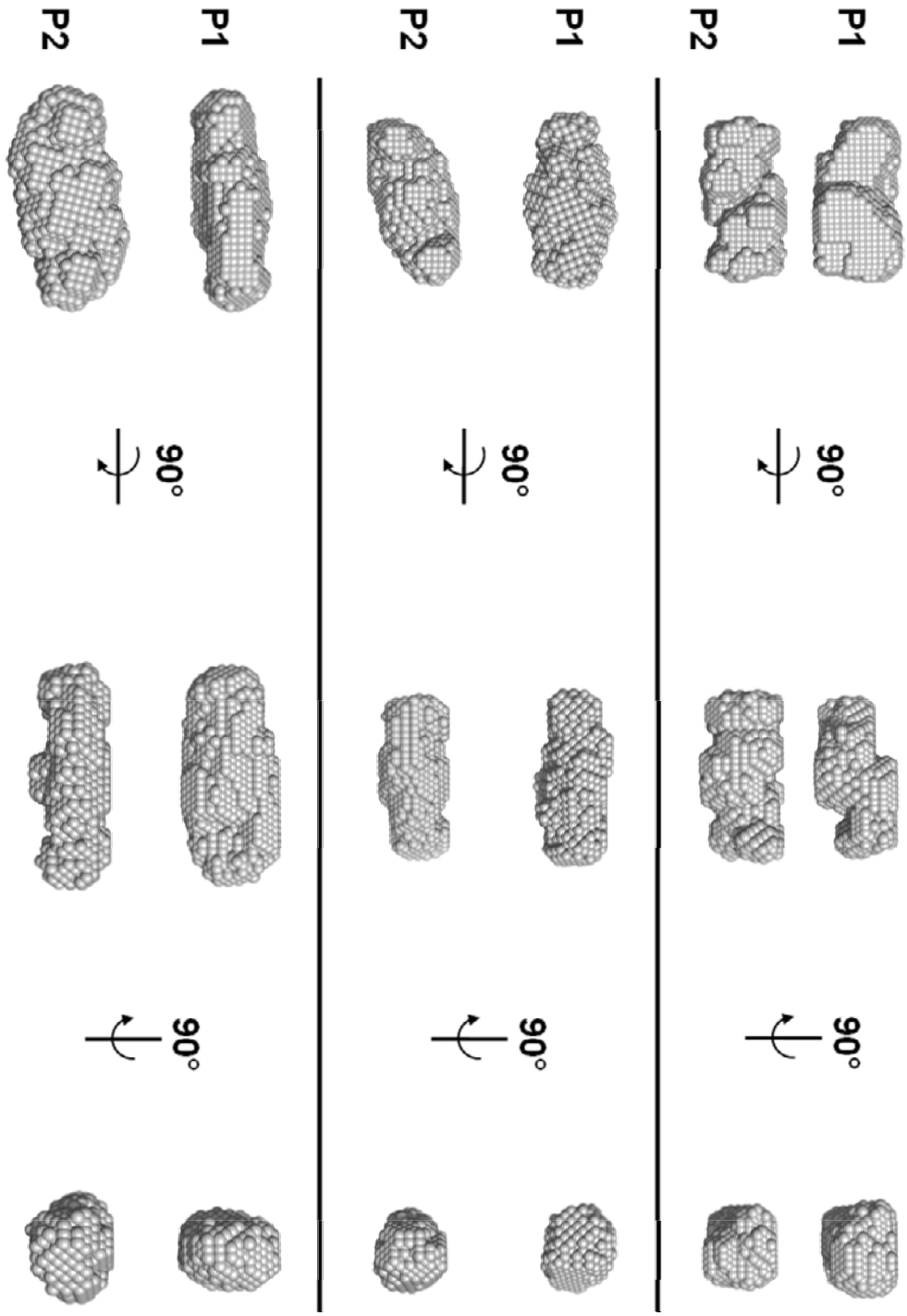

Figure 2.3.3 Averaged Low Resolution *ab initio* Shape Reconstructions for WC I/II (705-918)/(174-455). Averaged consensus envelopes for 0.8, 1.7 and 3.3 mg/mL (top, middle and bottom) show a slight increase in dimensions with increased concentration. Also observed is a basic retention of symmetrical features for models with and without a P2 symmetry constraint.

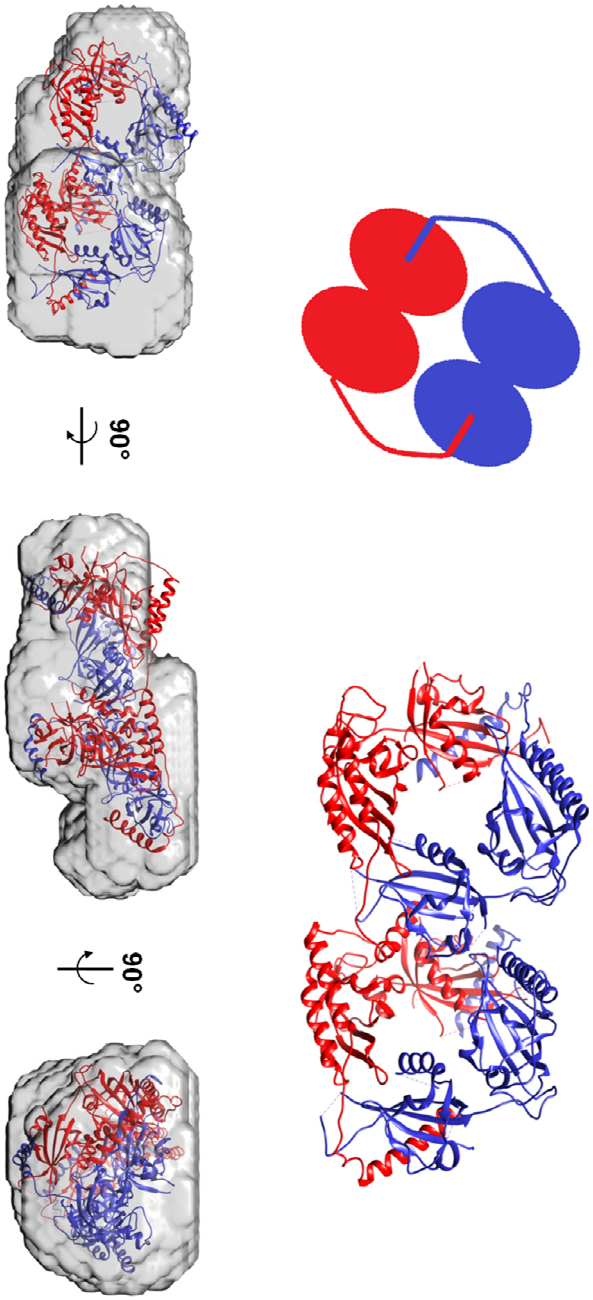

Figure 2.3.3.1 Superposition of a Canonical PAS Structure and the WC I/II (705-918)/(174-455) 0.8 mg/mL Envelope. An cartoon showing the Drosophila Period PAS AB dimmer orientation (top left) and corresponding ribbon model (PDB: 3RTY) of two PAS AB dimmers (top right) used for superposition with the WC I/II (705-918)/(174-455) 0.8 mg/mL P1 symmetry constrained consensus envelope (bottom). From Chimera fitting, good agreement is observed between the low resolution envelope and the ribbon model, suggesting the low resolution envelope has a calculated volume sufficient to accommodate two WC I/II PAS dimers

918)/(174-455) 0.8 mg/mL envelope and two Period PAS structures, indicating the low resolution envelopes have a calculated volume sufficiently to accommodate two WC I/II PAS dimers.

2.3.6 WC I/II (705-918)/(174-455) SEC-MALS Analysis

To accurately determine the molecular weight of the WC I/II sample, SEC was used in conjunction with MALS. The experimental setup consisted of SEC separation in a continuous flow experiment with a DAWN HELEOS-II and Optilab T-rEX (Wyatt Technology) for mass and concentration determination respectively. Analysis of a sample concentrated to 6.5 mg/mL using ASTRA software (Wyatt Technology) and Igor Pro (WaveMetrics), indicates the presence of a 118 (±4%) kDa species (Figure 2.3.3.2). The observed light scattering and RI signal for the tested sample displays a relative low intensity, resulting in a 4% error in the mass measurement. This effect was due to sample storage for 15 min at room temperature prior to injection, resulting in the formation of a visible white-precipitate, which was removed with a 13,000 rpm centrifugation at 4 °C for 10 min. Upon completion of centrifugation, sample concentration was not obtained as the sample was immediately subjected to SEC-MALS analysis.

2.3.7 WC I/II (705-1000/179-500) – Addition of the Zinc fingers

Having optimized protein handling steps for WC I/II (705-918)/(174-455), efforts were turned to White Collar I and II proteins containing the C-terminal zinc

finger domains. Results for WC I/II (705-1000) /(179-500) experiments are concisely presented. For a thorough treatment of experimental SAXS considerations, refer to corresponding WC I (705-918) discussion.

2.3.7.1 WC I/II (705-1000/179-500) SEC and SDS-PAGE Analysis

Purification of WC I/II (705-1000)/(179-500) was performed with IMAC followed by SEC. The principle elution peak from SEC eluted at approximately 179 mL (Figure 2.3.3.3), suggestive of a species with an apparent molecular weight of approximately 131 kDa, smaller than the predicted theoretical molecular weight of 146 kDa for a WC I/II complex. This elution value differs substantially from previous purification of individual WC I (705-1000) and WC II (179-500) proteins with SEC elutions of 204 and 193 mL respectively as shown in Chapter 1 Figures 1.3.2.3 and 1.3.1.3. Additional analysis of the SEC profile indicates a small shoulder eluting near 218 kDa.

SDS-PAGE gel analysis of SEC elution fractions (Figure 2.3.3.3) show minimal formation of high molecular weight bands. Low molecular weight bands resulting from apparent loss of WC I was readily separable as seen in protein concentrated to 2.3 mg/mL. Bands with molecular weights consistent with the overexpressed WC I/II are present in each of the fractions collected for concentration as denoted by red lines and in the concentrated protein.

Quantification of relative band intensity with Syngene Genetools indicated

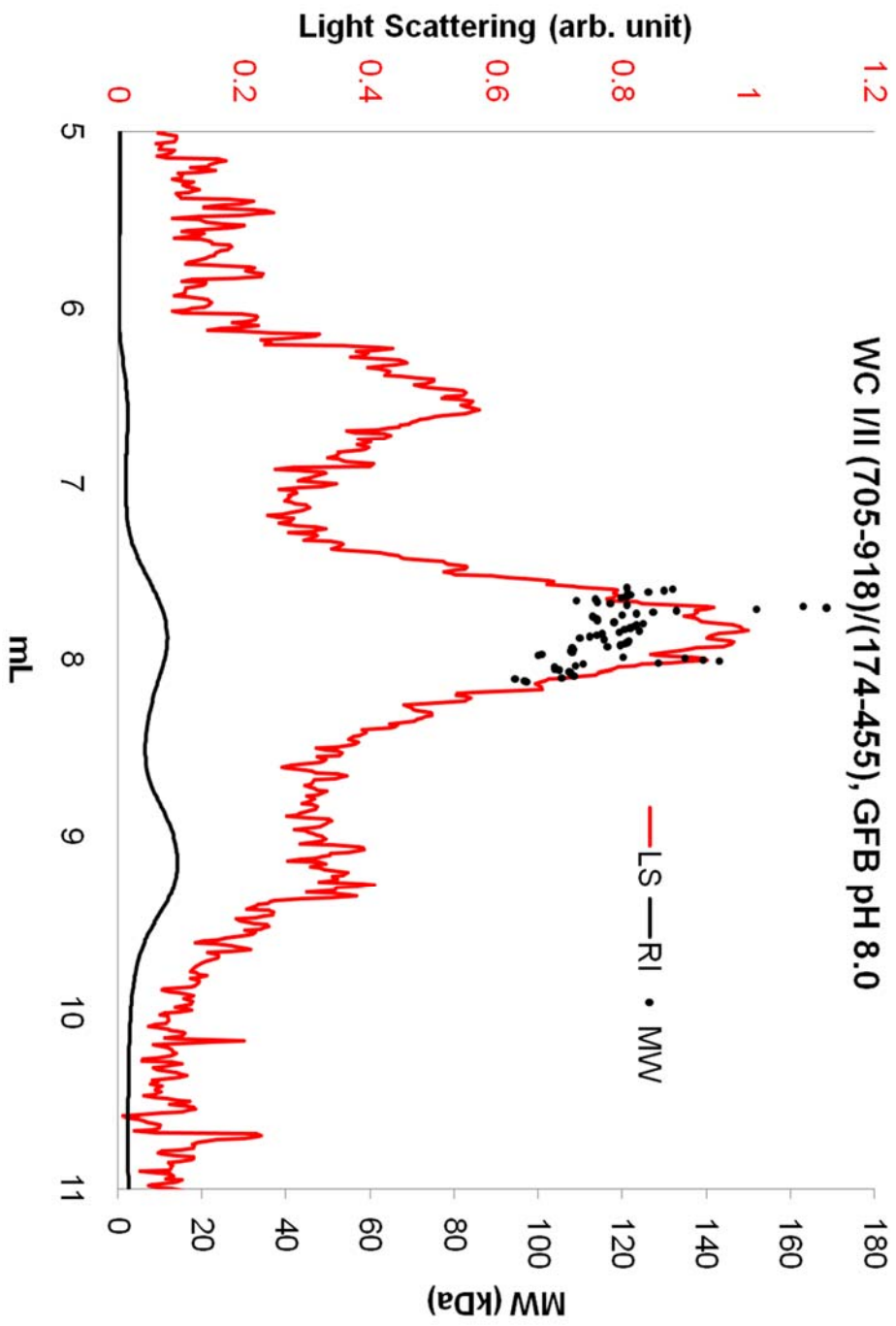

Figure 2.3.3.2 Size Exclusion Chromatography Coupled with Multi-angle Light Scattering for WC I/II (705-918)/(174-455). Static light scattering for WC I/II (705-918)/(174-455) concentrated to 6.5 mg/mL indicates an absolute molecular weight of 118 (±4%) kDa, closely matching the theoretical molecular weight for a WC I/II complex. Low light scattering and refractive index signals are due to low sample concentration due to poor solubility at room temperature.

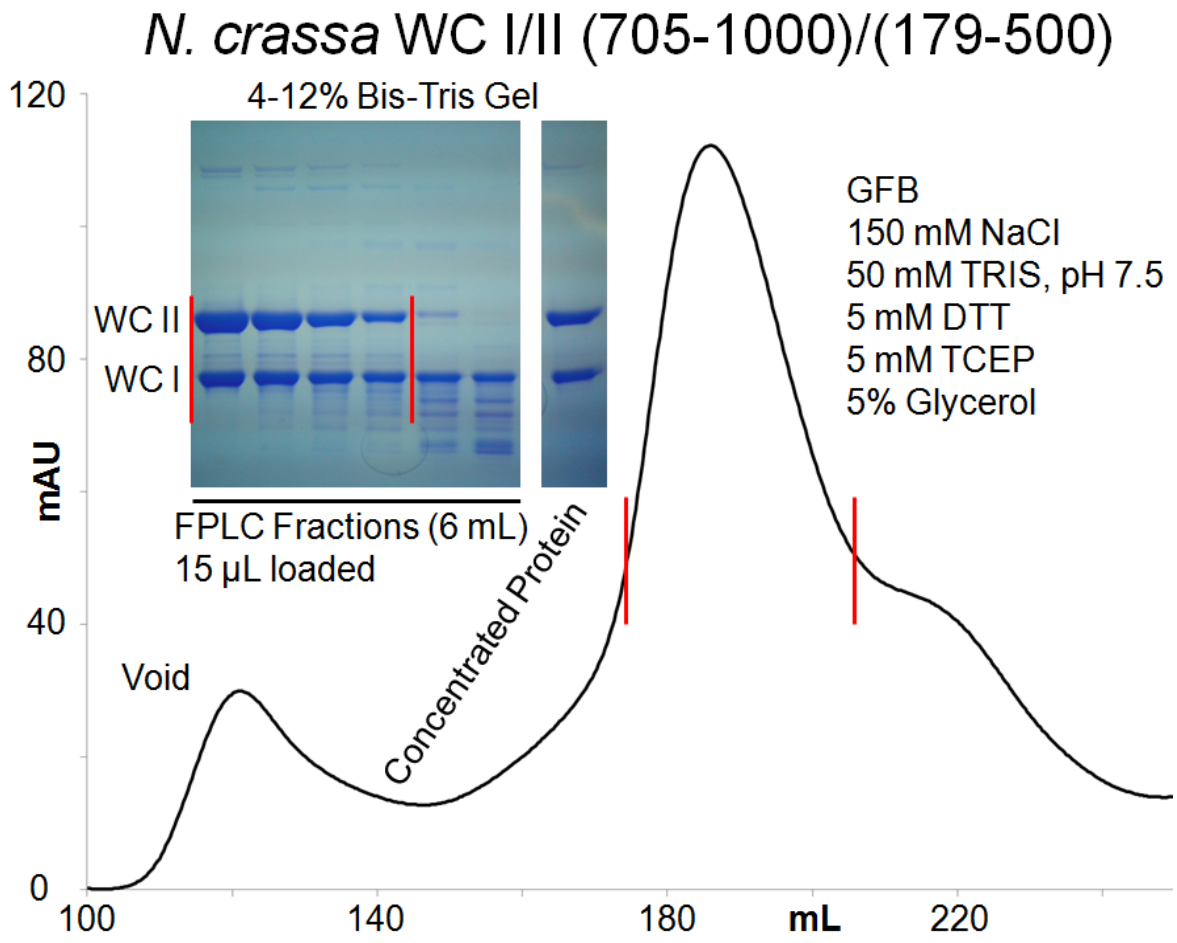

Figure 2.3.3.3 Copurification of WC I/II (705-1000)/(179-500). SDS-PAGE gel analysis of SEC elution fractions (upper left), shows protein corresponding to WC I (705-1000) and WC II (179-500) theoretical molecular weight. Protein concentrated to 2.3 mg/mL shows an absence of high and low molecular weight bands. The SEC elution profile has a principle elution peak at 179 mL, suggestive of a species with a 131 kDa apparent molecular weight, smaller than the predicted 146 kDa molecular weight associated with a WC I/II tetrameric complex. Red bars on elution profile correspond to the SEC elution fractions selected for concentration.

an excess of WC II for the early portion of the primary elution peak transitioning to an excess of WC I in the later elution volume. Further, SEC fractions selected for concentration show an approximate excess of 20% for WC II relative to WC I in the concentrated protein sample (Table 2.3.2.1).

Table 2.3.2.1 SDS-PAGE Gel Analysis of WC I/II (705-1000/179-500) SEC Fractions Selected for Concentration

	Lane					
Protein	**1 (%)**	**2 (%)**	**3 (%)**	**4 (%)**	**Flow Through (%)**	**Concentrated Protein (%)**
WC II	62	58	55	48	67	55
WC I	38	42	45	52	33	45

2.3.8 *N. crassa* WC I/II (705-1000)/(179-500) SAXS Analysis

2.3.8.1 Evaluating WC I/II (705-1000)/(179-500) for Radiation Damage

Buffers and samples were monitored for radiation damage by checking 15 sequential 30 s I(q) versus q plots with linear X–linear Y and log X–log Y axis for shifts in q values. As seen in Figure 2.3.3.4, curves generated using RAW overlap with no shift in q values, indicating the samples did not undergo aggregation or radiation damage while exposed to the beam [24].

2.3.8.2 Generating WC I/II (705-1000)/(179-500) Interference-free Scattering Curves

To determine if the WC I/II (705-1000)/(179-500) samples had negligible interparticle effects and could represent infinitely dilute approximations, samples were first averaged, then buffer subtracted averages of scattering data were generated and superpositioned with linear X–log Y axis using RAW. From data analysis (Figure 2.3.3.5), no substantial change in q values is observed in the tested samples indicating samples can be considered infinitely dilute.

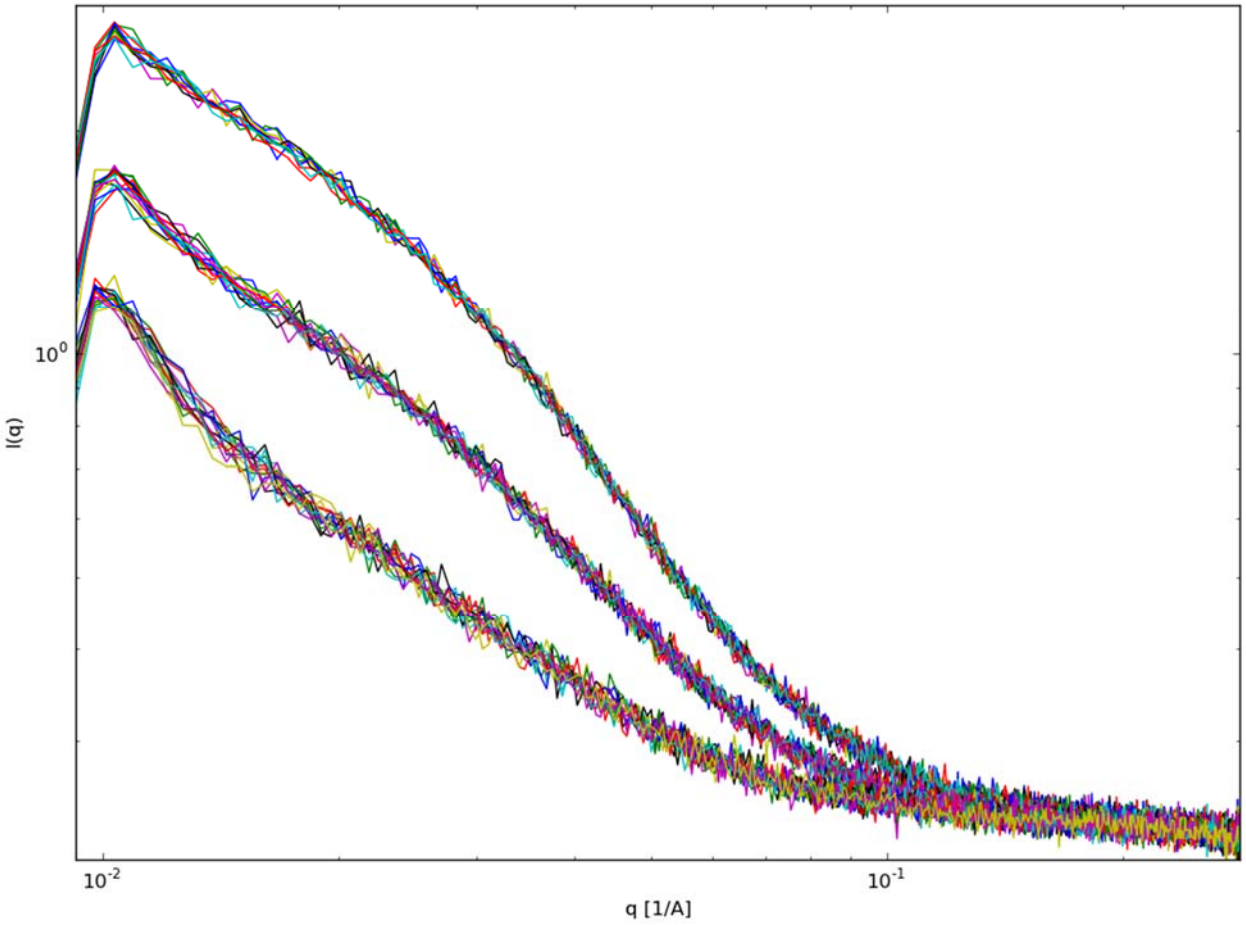

Figure 2.3.3.4 Scattering Profiles for WC I/II (705-1000)/(179-500) Dilutions. Log X–log Y plots for WC I/II (705-1000)/(179-500) at 2.3, 1.2 and 0.7 mg/mL (top, middle, bottom) show no change in q value, indicating protein samples did not undergo aggregation or radiation damage from beam exposure.

2.3.8.3 Determining WC I/II (705-1000)/(179-500) Rg and I(0) with Guinier Plots

Guinier analysis of scattering data was performed to evaluate data quality (Figure 2.3.3.6 and Table 2.3.2.2). As seen in Guinier plots generated with RAW, a linear fit to the scattering data with no up or downturns in intensity at low q values is observed suggesting the samples are free of aggregation and interparticle inference.

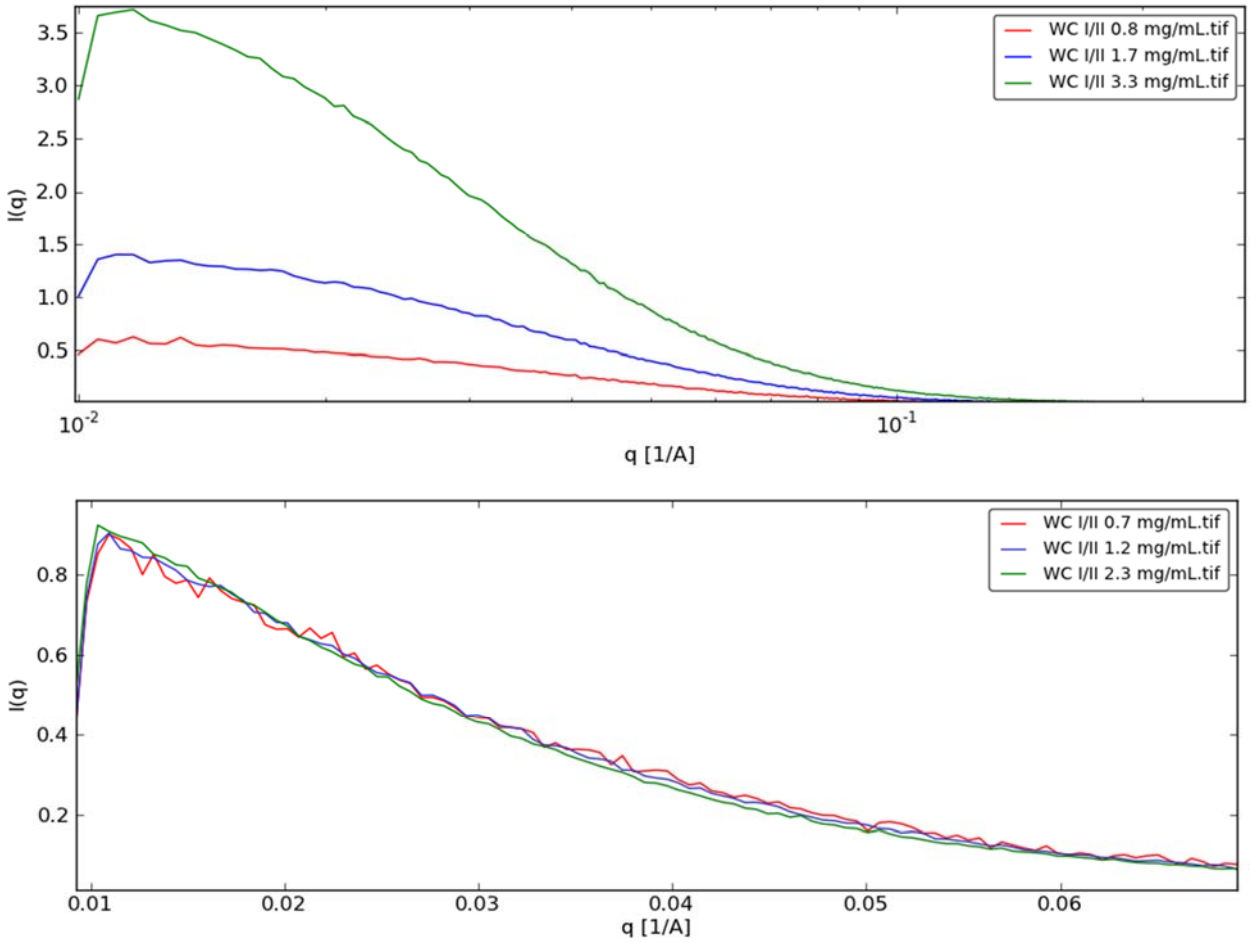

Figure 2.3.3.5 Superpositioned WC I/II (705-1000)/(179-500) Scattering Profiles. Averaged and buffer subtracted scattering profiles show relative I(q) for tested samples (top panel). Superpositioned profiles (lower panel) do not indicate the occurrence of interparticle interactions and samples at 0.7, 1.2 and 2.3 mg/mL can be considered as infinitely dilute.

Table 2.3.2.2 Values for WC I/II (705-1000)/(179-500) RAW Guinier Plots

Concentration	qStart	qEnd	qRg	Rg	I0	rsq
0.7 mg/mL	0.0115	0.02648	1.30162	50.24288	1.63547	0.95084
1.2 mg/mL	0.0109	0.02533	1.30934	52.89146	1.86207	0.9938
2.3 mg/mL	0.0104	0.02418	1.31879	55.87177	2.07716	0.97965

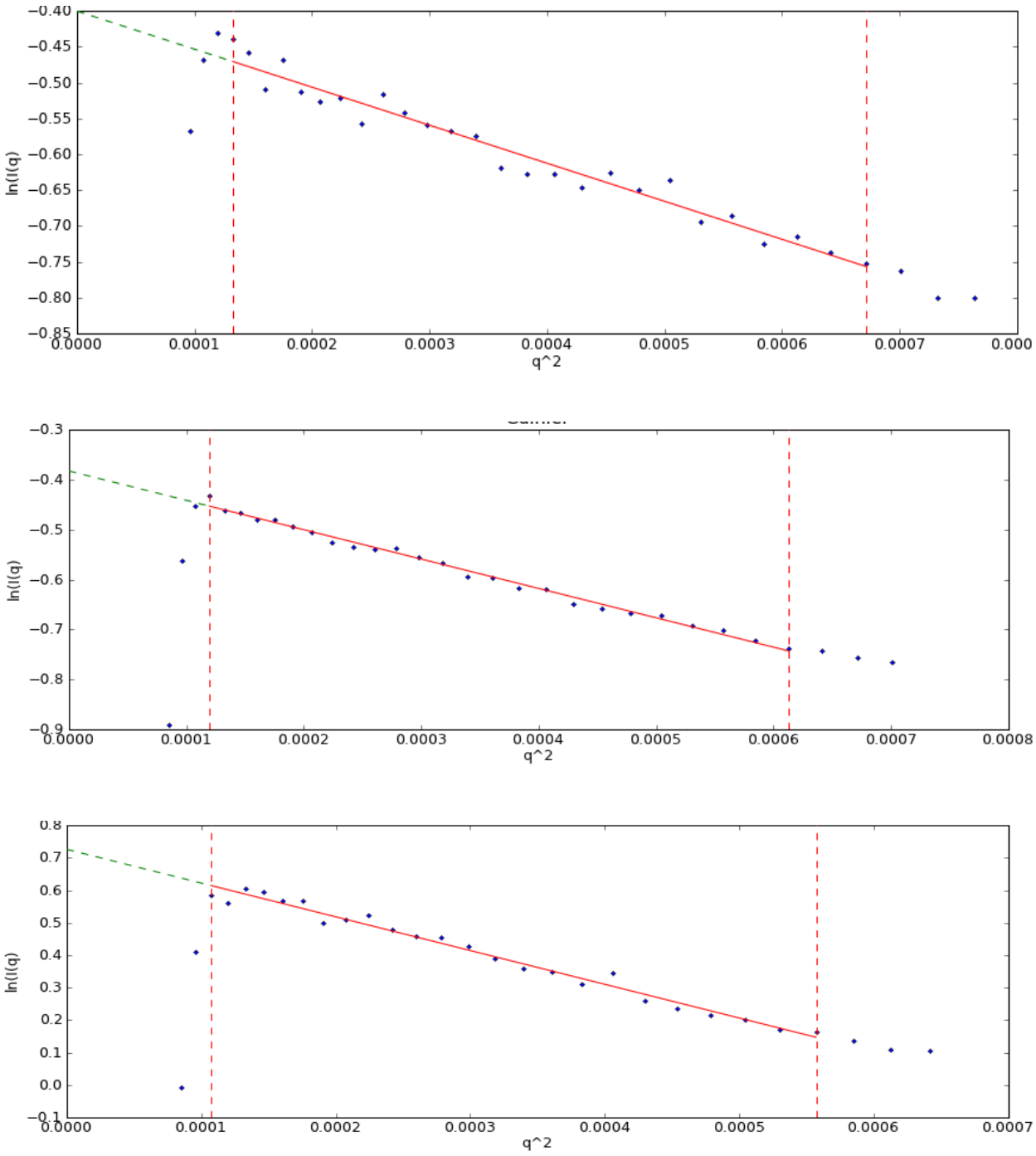

Figure 2.3.3.6 Guinier plots for WC I/II (705-1000)/(179-500). A linear fit is obtained in the Guinier region for samples at 0.7, 1.2 and 2.3 mg/mL (top, middle and bottom), with no change in intensity at low q values, suggestive of non-aggregated samples.

2.3.8.4 Evaluating WC I/II (705-1000)/(179-500) Scattering Data with PRIMUS

PRIMUS was used to obtain estimates of data quality, P(r), Dmax, Porod volume as well as Rg and I(0) (Table 2.3.2.3). Using the P(r) function provided by AUTORG, I(0) and Rg estimates were generated. Rg values show good agreement with those obtained from RAW, however, I(0) values for 0.7 and 1.2 mg/mL WC I/II (705-1000)/(179-500) differ significantly, while 2.3 mg/mL has good agreement (Table 2.3.2.2 and Table 2.3.2.3). GNOM estimates of data quality were reported above 70% (good) and aggregation was not indicated for tested samples.

P(r) functions for WC I/II (705-1000)/(179-500) at 0.7,1.2 and 2.3 mg/mL (Figure 2.3.3.7) show a smooth slope near Dmax, lacking bumps which are indicative of protein aggregation.

Table 2.3.2.3 AUTORG and GNOM Estimates for WC I/II (705-1000)/(179-500)

	0.7 mg/mL	1.2 mg/mL	2.3 mg/mL
AUTORG			
Rg	49.025 ± 4.775	52.585 ± 1.1221	56.803 ± 0.442
I(0)	0.375 ± 0.006	0.981 ± 0.006	2.083 ± 0.007
SRg Limits	0.677 (1.298)	0.605 (1.302)	0.589 (1.243)
Fidelity	0.793	0.901	0.924
Range	8-30	4-27	2-22
GNOM			
Quality Estimate [%]	86.9	76.7	77.2
Reciprocal Space Rg/I(0)	50.19 ± 0.35	55.93 ± 1.0	58.94 ± 2.1
Real Space Rg/I(0)	50.36 ± 0.35	56.24 ± 1.0	59.28 ± 2.1
Porod Volume	287537	324454	371590
Range	8-472	4-472	2-472
Dmax	157.95	196.61	206.93

2.3.8.5 WC I/II (705-1000)/(179-500) Molecular Weight Estimates

Molecular weight estimates for WC I/II (705-1000)/(179-500) samples were generated from comparison of forward scattering data to that of known standards, SAXSMOW and analysis of the Porod volume for comparison (Table 2.3.2.4). As seen in Figure 2.3.3.8, SAXSMOW analysis of GNOM output shows significant noise due to low sample concentration, indicating the need for a qmax= 0.15 cutoff to minimize error in mass predictions.

Table 2.3.2.4 Molecular Weight Estimates for WC I/II (705-1000)/(179-500)

	I(0) (kDa)		SAXSMOW (kDa)			Vp (kDa)
Concentration	Lysozyme	Glucose Isomerase	0.25q	0.2q	0.15q	(Vp*1.2/2)
0.7 mg/mL	76	82	169	184	203	173
1.2 mg/mL	120	128	193	210	234	195
2.3 mg/mL	132	141	215	235	261	223

Molecular weight estimates derived from analysis of the Porod volumes as calculated by GNOM using a relationship of (Vp*1.2)/2= molecular weight, following an increasing trend with concentration similar to values obtained from SAXSMOW. However, values obtained from the comparison of sample I(0) values to those of lysozyme and glucose isomerase indicate a sharp rise in molecular weight with increasing from 0.7 to 1.2 mg/mL and then a smaller increase to 2.3 mg/mL (Figure 2.3.3.9).

2.3.8.6 Concentration Dependent Increases in WC I/II (705-1000)/(179-500) SAXS Data

While analysis of scattering data with superposition, Guinier and P(r) plots

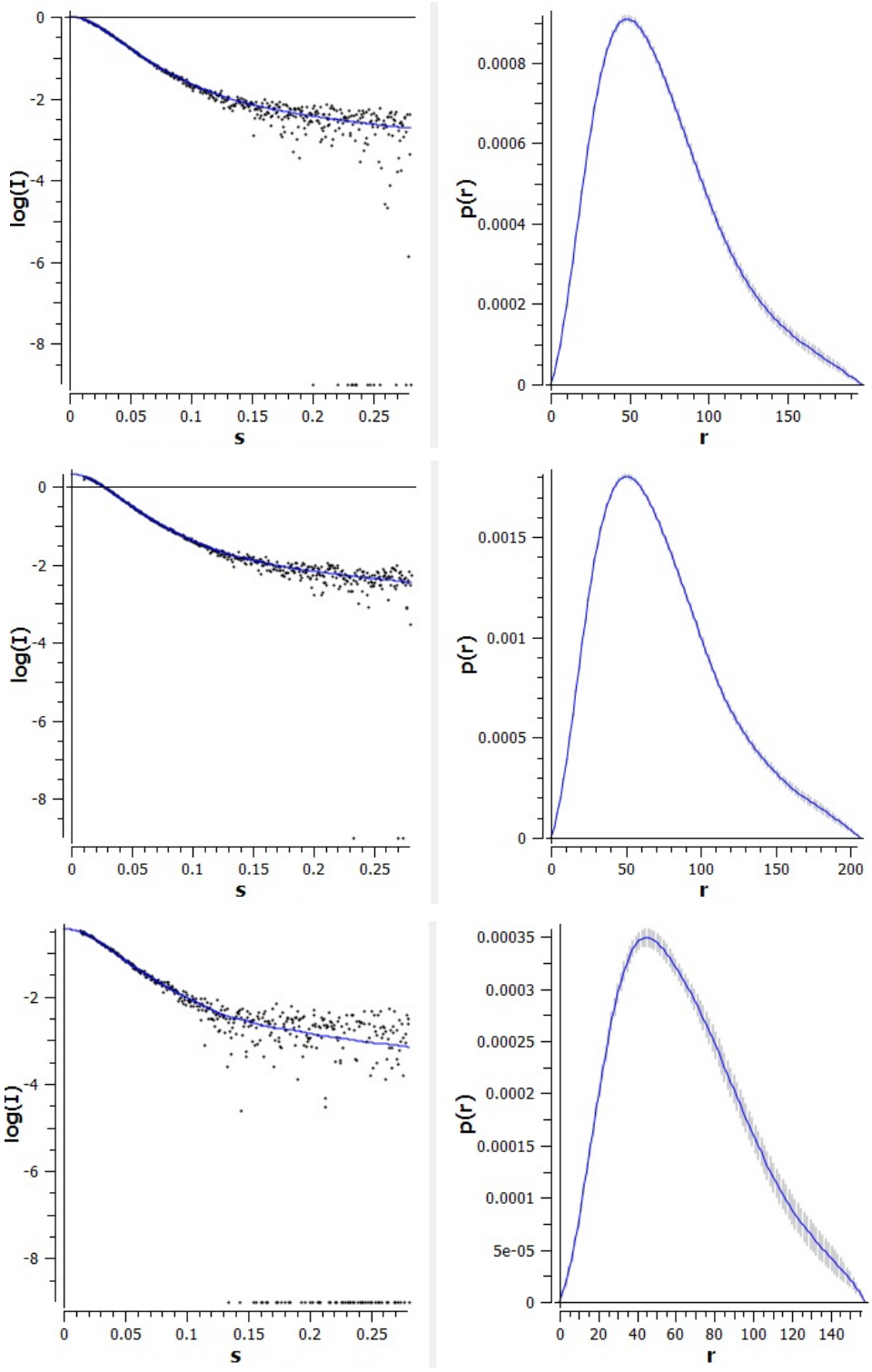

Figure 2.3.3.7 P(r) Plots for WC I/II (705-1000)/(179-500). P(r) functions (with fit on left) for WC I/II (705-1000)/(179-500) samples at 0.7, 1.2 and 2.3 mg/mL (top, middle and bottom) show a smooth slope near Dmax, indicative of non-aggregated samples.

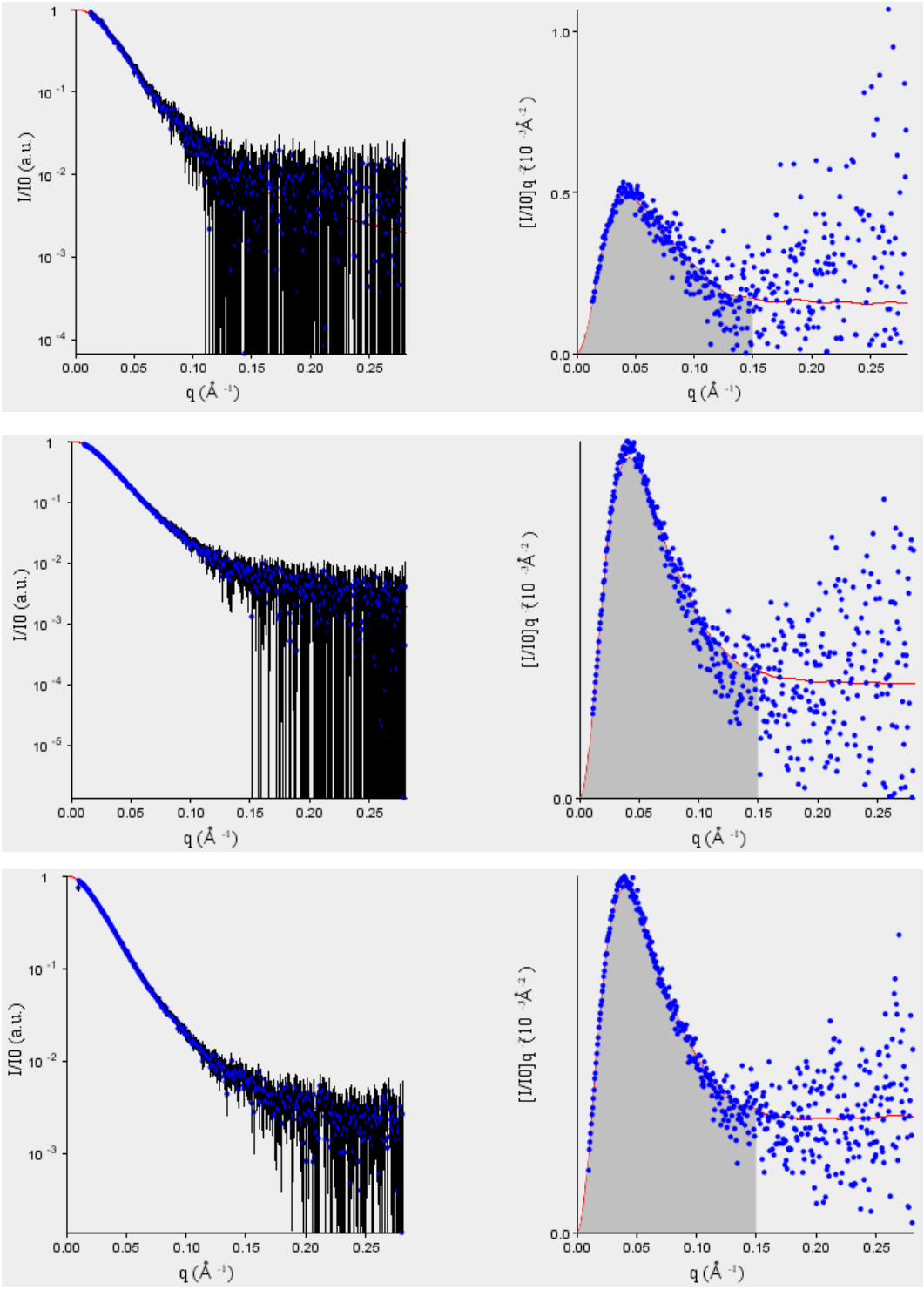

Figure 2.3.3.8 SAXSMOW Analysis of WC I/II (705-1000)/(179-500). SAXSMOW analysis of Porod regions for 0.7, 1.2 and 2.3 mg/mL (top, middle, bottom) show a low signal to noise ratio, indicating the need for a qmax= 0.15 cutoff value in molecular weight estimates.

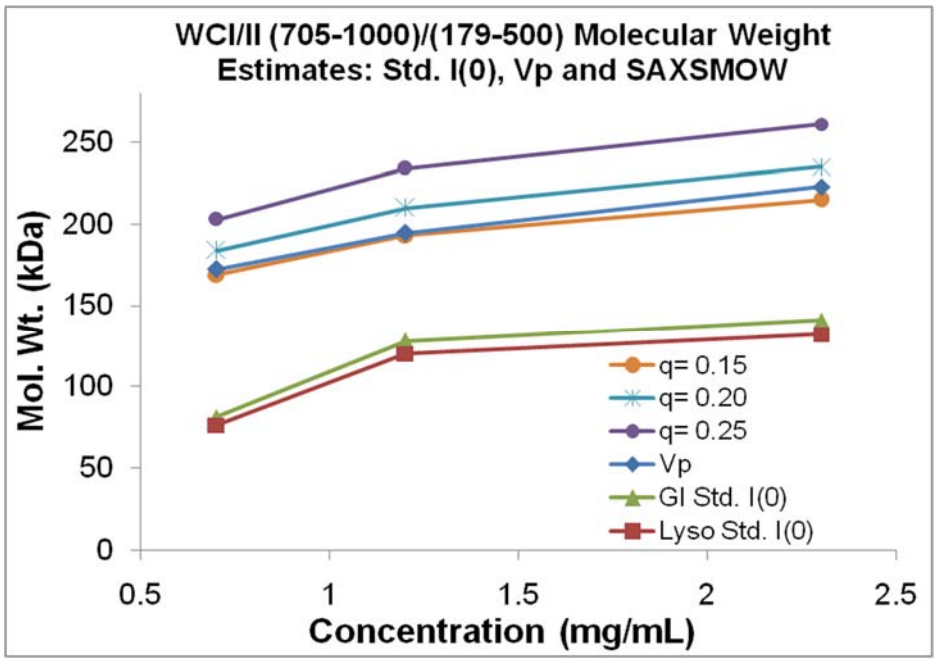

Figure 2.3.3.9 Molecular Weight Estimates for WC I/II (705-1000)/(179-500). Comparison of WC I molecular weight estimates from 1) I(0) values to known concentrations of lysozyme (Lyso Std.) and glucose isomerase standards (GI Std.), 2) SAXSMOW estimates based on variation in qmax (qmax= 0.15, 0.20 and 0.25) and 3) analysis of the Porod volume (Vp) as determined by PRIMUS using the relationship (Vp*1.2)/2= Mol. Wt., show a concentration dependent increase in predicted molecular weight.

did not indicate the presence of aggregation, a concentration dependent increase in I(0), Rg, Dmax, Vp (Table 2.3.2.2 and Table 2.3.2.3) and molecular weight estimates (Figure 2.3.3.9) was observed. Comparison of Rg and I(0) values obtained using RAW and PRIMUS show excellent agreement between estimated Rg values and I(0) values (Figure 2.3.4) with both showing a concentration dependent increase, suggestive of either a change in multimerization or aggregation.

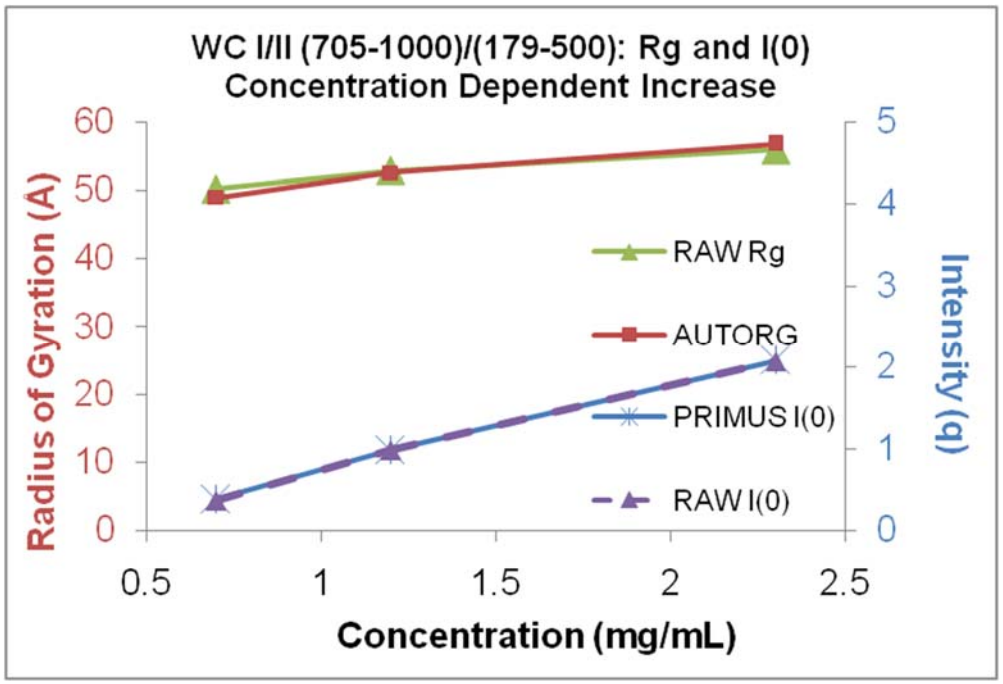

Figure 2.3.4 I(0) and Rg Concentration Dependent Increases in WC I/II (705-1000)/(179-500) Samples. Comparison of calculated Rg and I(0) values show a concentration dependent increase for samples at 0.7, 1.2 and 2.3 mg/mL, which is indicative of a change in multimeric state or concentration dependent aggregation.

2.3.8.7 Evaluating WC I/II (705-1000)/(179-500) Folding with Kratky Plots

A Kratky plot of WC I/II (705-1000)/(179-500) samples was generated to evaluate folding and compactness (Figure 2.3.4.1). From this plot a broad bell-shaped curve with no increase in $I(q)q^2$ with q is observed for sample at 2.3 mg/mL, indicative of a folded globular protein with an extended conformation. Samples at 0.7 and 1.2 mg/mL show a lower $I(q)q^2$ value, however, none of the concentrations tested indicate a plateauing or steady increase in $I(q)q^2$ with q, suggesting the WC I/II samples are folded.

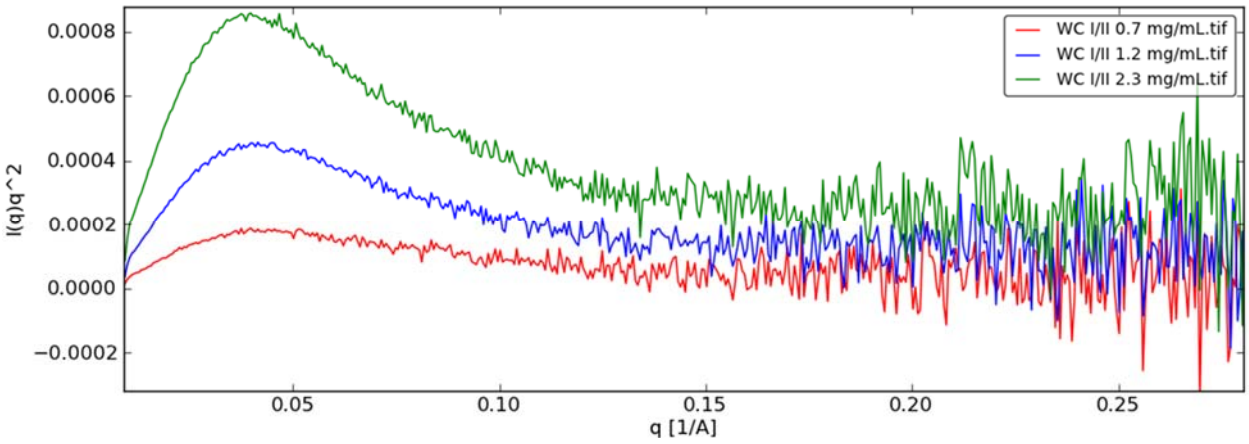

Figure 2.3.4.1 Kratky Plot for WC I/II (705-1000)/(179-500). Evaluation of scattering data shows the presence of a peak at low q and absence of a steady rise in $I(q)q^2$ with increasing q, consistent with a folded protein.

2.3.8.8 Generation of *ab initio* WC I/II (705-1000)/(179-500) Shape Reconstructions

For each concentration, DAMMIF was used to generate 10 low resolution three-dimensional *ab initio* shape reconstructions with and without a P2 symmetry constraint, except for the 1.2 mg/mL sample with P1 and 2.3 mg/mL with P2 symmetry constraints, which had 9 and 8 models generated respectively (not shown). Generated models were then evaluated using DAMAVER, which compared the set and determined the normalized spatial discrepancy for each of the low resolution reconstructions (not shown). Once determined, DAMAVER selected the model with the lowest NSD as the most probable model from which to generate a consensus envelope (Figure 2.3.4.2).

Average NSD values and their variations (ΔNSD) for WC I/II (705-1000)/(179-500) consensus envelopes indicate individual reconstructions have

greater spatial variation with the addition of the C-terminal zinc finger (Table 2.3.2.5).

Table 2.3.2.5 NSD Values for WC I/II (705-1000)/(179-500) Consensus Envelopes

	WC I/II (705-1000)/(179-500) Average NSD					
	0.8 mg/mL		1.7 mg/mL		3.3 mg/mL	
	P1	P2	P1	P2	P1	P2
Average NSD	1.126	0.846	0.981	1.033	0.887	0.844
ΔNSD	0.17	0.048	0.057	0.088	0.078	0.085

2.3.8.9 CRYSOL Fitting of Theoretical Mouse Period 2 Scattering Profiles to WC I/II Scattering Data

To evaluate whether the copurified WC I/II solution complexes are a close representation to a crystallographically determined canonical PAS structure, the WC I/II scattering data was compared to the theoretical scattering profile for mouse Period 2 PAS AB repeat (mPER2) (PDB: 3GDI), which as seen in Chapter 1 (Figure 1.4.1) has high sequence homology to the putative WC I/II PAS AB repeat. To facilitate this comparison, two mouse Period 2 PAS AB homodimers (WC I/II_tet) were arranged side by side using Xfit [34]. Also generated for comparison was a modified mPER2 structure, in which one of the WC I PAS A domains was deleted, resulting in an mPER2 PAS B/PAS AB structure (WC I/II_tetmod). This change reflects the presence of the single C-terminally extended WC I PAS B domain (705-918) instead of the PAS AB repeat (586-806), as shown in Figure 1.4.1. Finally, an octamer (WC I/II_oct) of the modified mPER2 PAS B/PAS AB structure was generated, as well as a 60%/40%

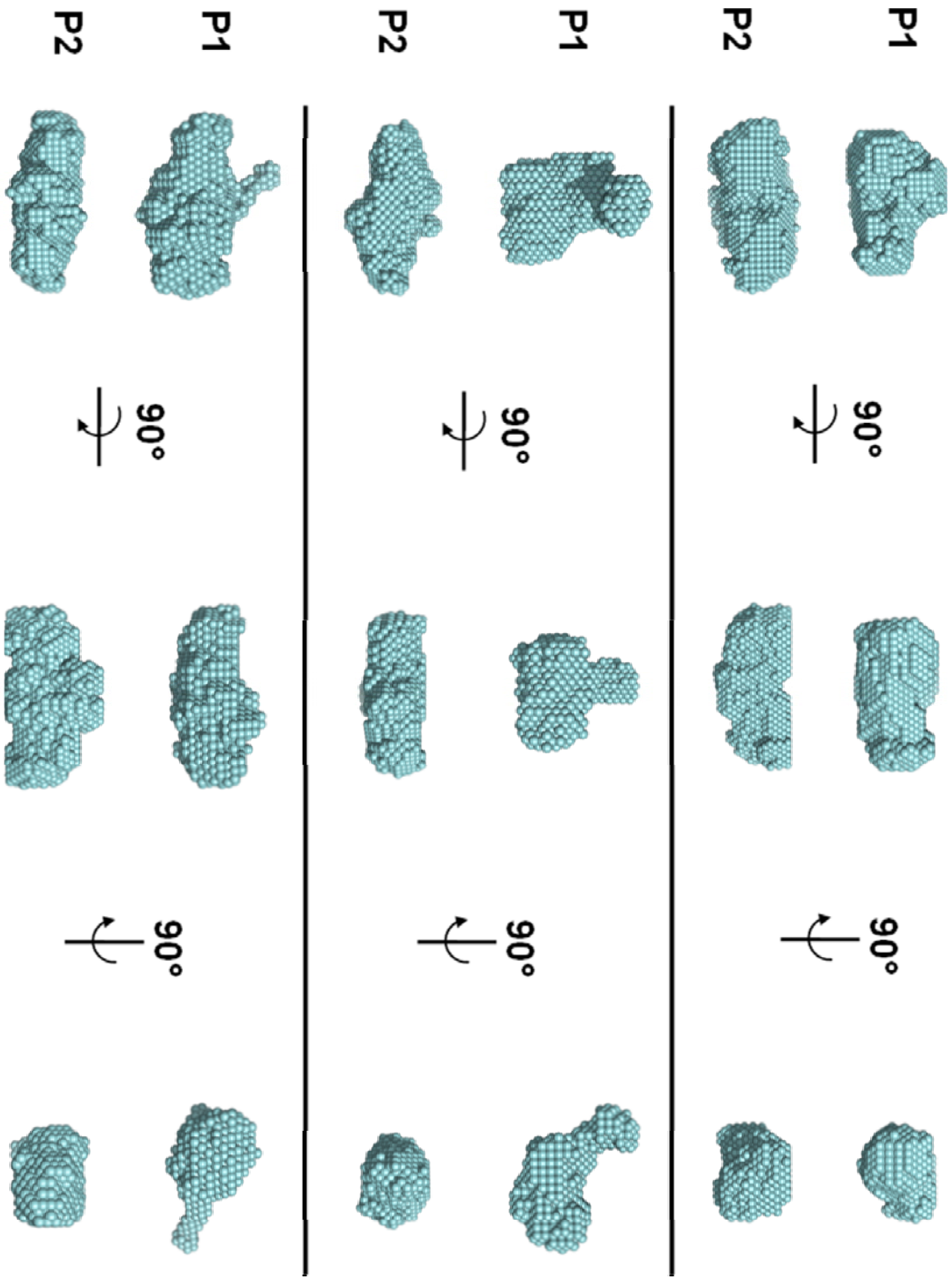

Figure 2.3.4.2 Averaged Low Resolution *ab initio* Shape Reconstructions for WC I/II (705-1000)/(179-500). Averaged consensus envelopes for 0.7, 1.2 and 2.3 mg/mL (top, middle and bottom) show a slight increase in dimensions with increased concentration. Also observed is a basic retention of symmetrical features for the 0.7 mg/mL models with and without a P2 symmetry constraint. A loss of symmetry is observed between the P1 and P2 models for the 1.2 and 2.3 mg/mL samples.

combination of modified mPER2 tetramer/octamer (WC I/II_octmod), which was assessed to account for possible aggregation in WC I/II scattering data.

To compare the different mPER2 models to the WCI/II solution complex, CRYSOL [23] was used to generate and fit theoretical scattering curves for the models to the WC I/II scattering data. As seen in Figure 2.4.1, WC I/II_tet, WC I/II_tetmod, WC I/II_oct, WC I/II_octmod (left panel) and WC I/II_Zntet, WC I/II_Zntetmod, WC I/II_Znoct, WC I/II_Znoctmod (right panel) overall show good agreement to both the WC I/II (705-918)/(174-455) at 0.8 and (705-1000)/(179-500) at 0.7 mg/mL for q values ($Å^{-1}$) ≤ 0.10 respectively, with the variation (χ^2) near 1.0 for each of models.

As aggregation is commonly seen when q values approach 0, inspection of the WC I/II_tet and WC I/II_tetmod fit at low q for proteins with and without the C-terminal zinc finger (left and right panel respectively) indicate the curves have a larger negative log(I) values than that of the WC I/II scattering data. With further assessment of the experimental fits showing WC I/II_tet has the greatest deviation at higher q values for both WC I/II scattering data sets. This contrasts the fit of the WC I/II_oct (and WC I/II_Znoct) at small q values, which is less negative than the experimental data. Finally, the fit of WC I/II_octmod (and WC I/II_Znoctmod) is the closet match to the experimental data, which further suggest the presence of aggregation in the WC I/II solution samples.

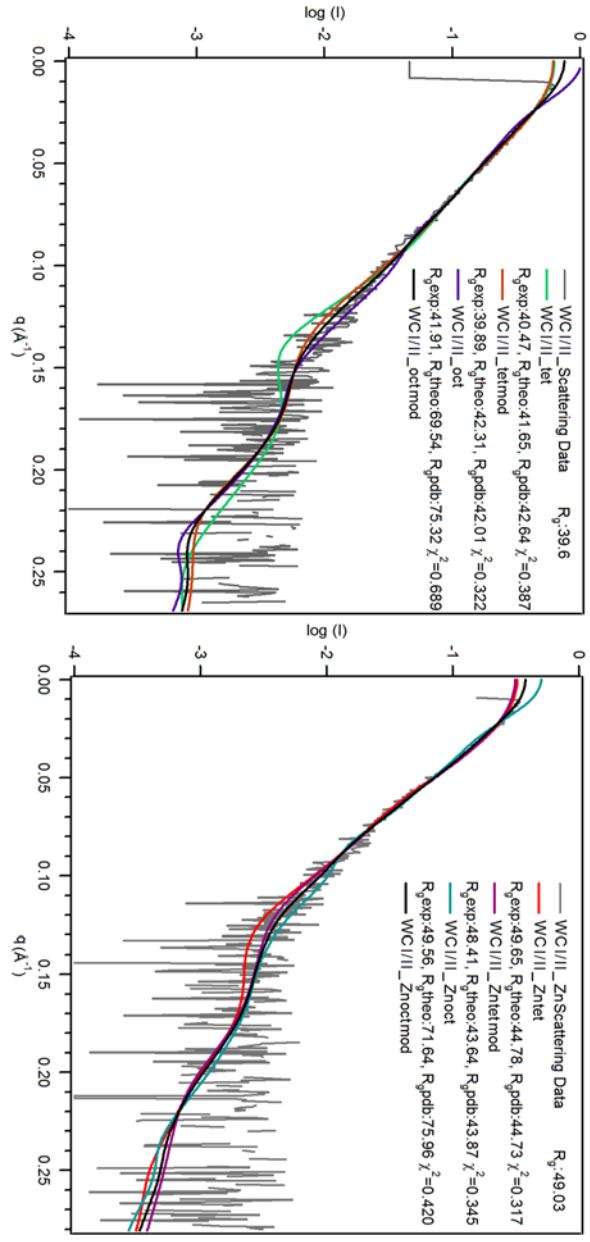

Figure 2.3.5 CRYSOL Fitting of mouse Period 2 PAS Repeat Models to *N. crassa* WC I/II Scattering Data. Theoretical scattering curves for mouse Period 2 PAS AB (WC I/II_tet), PAS B/PAS AB (WC I/II_tetmod), a linear octamer of PAS B/PAS AB (WC I/II_oct) and a 60%/40% combination tetramer/octamer (WC I/II_octmod) to approximate aggregation, as well as the corresponding models with C-terminal zinc fingers were fit to *N. crassa* WC I/II (705-918)/(174-455) at 0.8 mg/mL (left) and (705-1000)/(179-500) (right) scattering data at 0.7 mg/mL using CRYSOL. In both cases, the WC I/II_octmod (and WC I/II_Znoctmod) theoretical curves show the best fit at very low q angles, suggestive of aggregation in the WC I/II scattering data.

2.4 Discussion

From SDS-PAGE and SEC analysis, copurification of *N. crassa* WC I and II with and without their respective C-terminal zinc finger domains resulted in soluble, non-aggregated proteins. SEC elution volumes and apparent molecular weights for WC I/II (705-918)/(174-455) and (705-1000)/(179-500) were distinct from that of the individual proteins, which as shown in Chapter 1 eluted with apparent molecular weights consistent with dimers. This contrasted the primary SEC elution peak for both of the copurified proteins, where the hydrodynamic radius of approximately 130 kDa was estimated, suggestive of a WC I/II tetramer.

These observations are further supported by the absolute molecular weight determination of WC I/II (705-918)/(174-455) using SEC-MALS, which indicated a value of 118 (±4%) kDa, consistent with the predicted molecular weight of 119 kDa for a WC I/II (705-918)/(174-455) tetrameric complex.

Comparison of values obtained from SAXS experiments for the individually purified WC I and copurified WC I/II samples further support the conclusion that a WC I/II PAS domain mediated tetrameric complex was formed. As seen in reported values, Dmax and Rg for the individual WC I proteins were substantially smaller than those observed for the copurified WC I/II proteins. An increase in calculated Porod volumes was also observed, as the volumes for WC I proteins were only 40-45% of those observed for the copurified WC I/II proteins at the concentrations tested.

Additional analysis of WC I and WC I/II SAXS scattering data, showed subtle shifts in larger q values in the superposition plots as well as concentration dependent increases in I(0), Rg and Dmax, particularly for the most concentrated samples. These changes are suggestive of the onset of aggregation, which was also reflected as an in increase in the dimensions for the low resolution consensus envelopes.

While the accurate assessment of Kratky plots was complicated by low scattering intensity, the plots indicate the WC proteins display globular folding with an extended conformation, which is also reflected in the consensus envelopes as they consistently showed an extended globular structure. Another interesting feature estimated with the WC I (705-918) and WC I/II (705-918)/(174-455) consensus envelopes is a conservation of symmetry for envelopes generated with and without a P2 symmetry constraint, while the WC (705-1000)/(179-500) envelopes showed variations between the P1 and P2 constraints. One possibility for this difference could be attributed to the increased flexibility imparted by the proteins zinc finger domains.

Having demonstrated the theoretical scattering profiles for the mPER2 models closely fit the experimental WC I/II scattering data, and can be used to approximate the WC I/II solution complex, the WC I/II_tetmod (mPER2 tetramer), with and without the C-terminal zinc fingers were visualized with PYMOL. As seen in Figure 2.4.1, the models reflect the elongated shape observed in the WC

I/II (705-918)/(174-455) and (705-1000)/(179-500) consensus envelopes. In this model, the four zinc fingers, as represented by blue spheres, are aligned in a row.

As the models were shown to closely fit the scattering data and match the basic dimensions seen in the low-resolution envelopes, this data further supports the formation of a tetrameric WC I/II solution complex observed in this study. This conclusion also suggests that *N. crassa* WCC is a tetrameric species, which is formed from interactions between the putative PAS AB repeat indicated from in the comparison of the WC I and II sequences to *M. musculus* and *D. melanogaster* PERIOD as well as *M. musculus* CLOCK and BMAL1 in Chapter 1.

To evaluate possible binding of the WC I/II_tet model to separate cis-acting proximal and distal light-acting response elements in the *frq* promoter region, first identified as conserved imperfect GATA (GATN) repeats [7, 8] and in later work as pseudo-palindromic repeats [9], the WC I/II_tet was aligned to a known structure of zinc fingers bound to DNA. In this model, the zinc fingers are spaced to accommodate the distance imparted from the PAS domains. As seen in Figure 2.4.2 A, the DNA bound zinc fingers (shown as blue spheres) align well to the PAS domains, suggesting the linear arrangement of zinc fingers in the tetrameric model could simultaneously bind multiple DNA sequences.

Interestingly, the identified binding sequences in *frq* promoter region as well as an identified region for submerged protoperithecia-1 (SUB-1), a GATA

family transcription factor whose expression is under control of the WCC and is involved in fine-tuning the organisms light response, vary in the number of conserved binding sequences (Figure 2.4.2 B). This suggests that oligomeric state of the complex might change throughout the circadian day, as the number of available binding sites change.

More work will be needed to help elucidate the structure of this assembly and determine if the orientation of the C-terminal zinc fingers, as represented in Figure 2.4.1 as large blue spheres, are oriented together in the complex or are in a flipped orientation. If the arrangement of the PAS domains as shown in Figure 2.4.1 is correct and the zinc fingers are aligned when binding the four pseudo-palindromic light response elements identified in the frequency promoter region (Figure 2.4.2), the complex would show similarity to the orientation of the basic helix-turn-helix (bHLH) domains in the mouse CLOCK:BMAL1 structure (PDB: 4F3L), which are seen in close proximity.

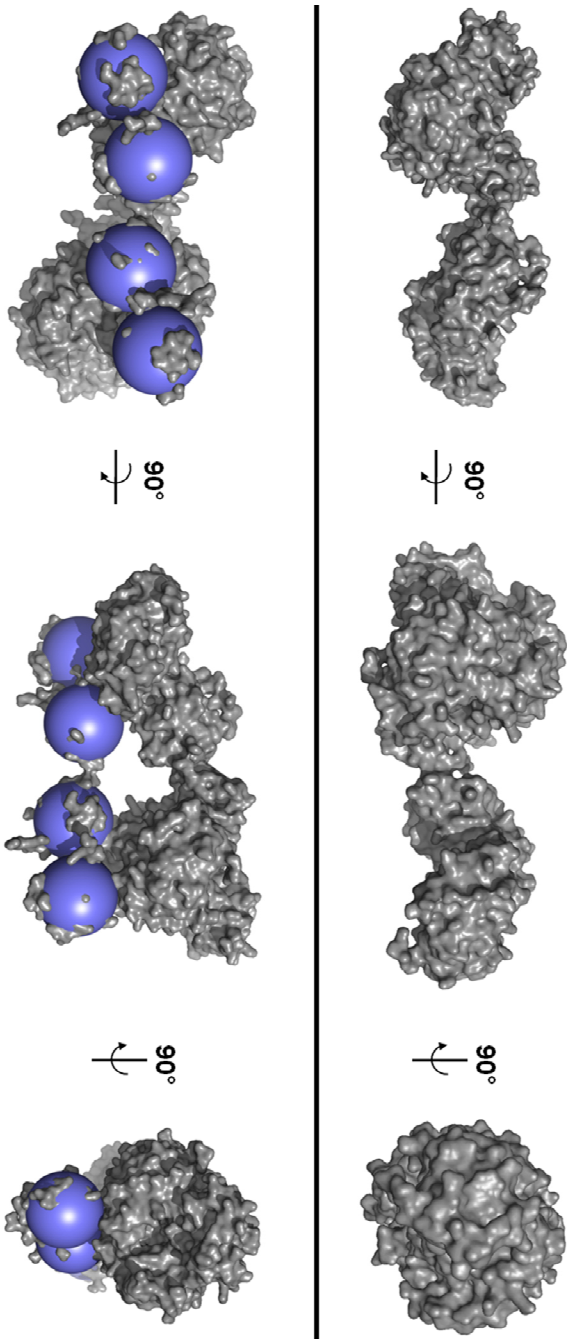

Figure 2.4.1 Orientation of PAS Domains in Mouse Period 2 PAS AB Repeat Model Suggest Aligned Zinc finger Domains in the *N. crassa* White Collar I/II Tetrameric Complex. Two mouse Period 2 PAS AB dimers (PDB: 3GDI) were arranged to form a PAS AB tetramer using Xfit (WC I/II_tet) for comparison to the WC I/II scattering data and low resolution consensus envelopes. The WC I/II_tet model shows conservation of the basic elongated structure as observed in *N. crassa* WC I/II (705-918)/(174-455) and (705-1000)/(179-500) envelopes.

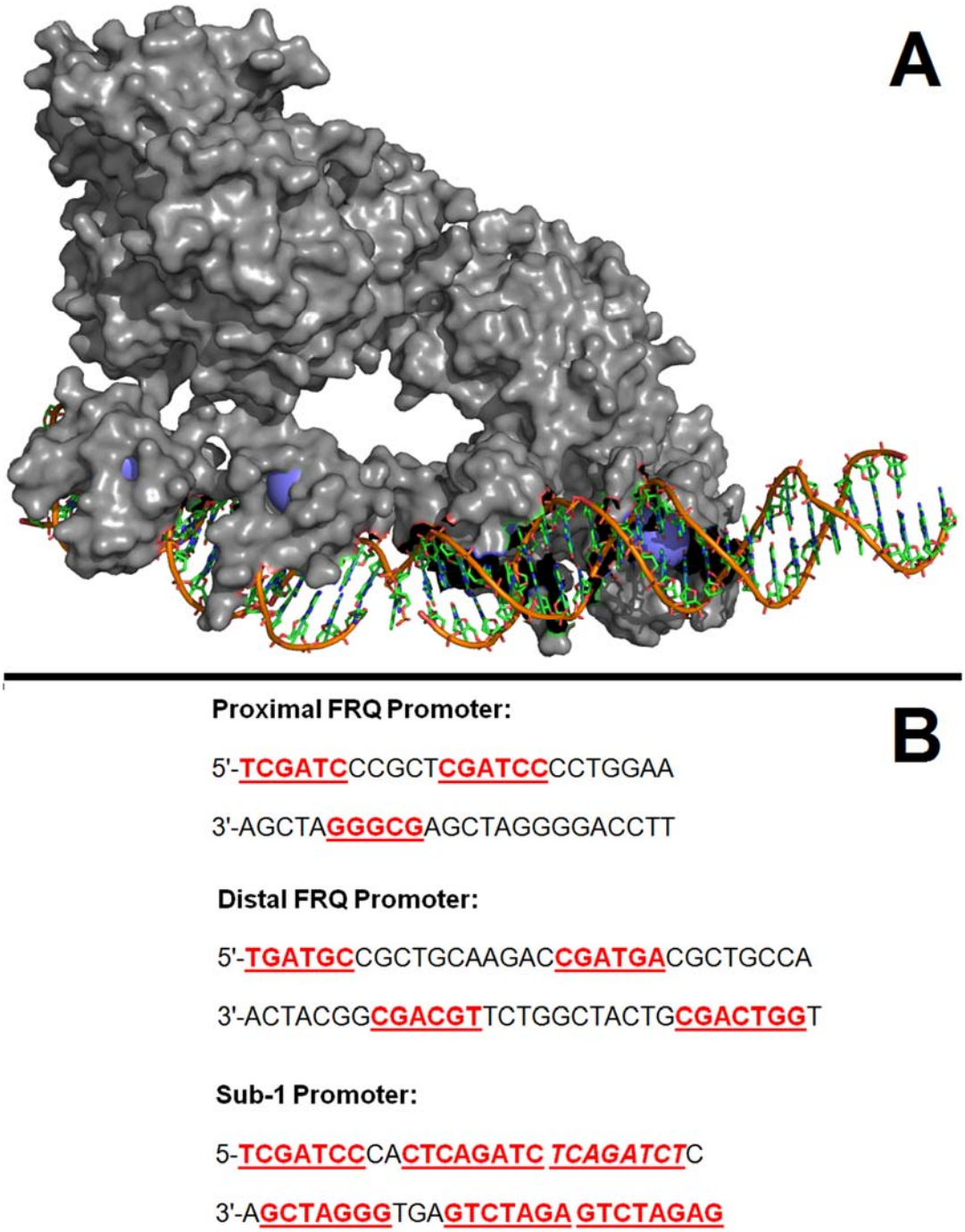

Figure 2.4.2 ***N. crassa*** **White Collar I/II Tetrameric PAS Complex DNA Binding Model.** A) Alignment of the WC I/II_tet model to a known structure of zinc fingers bound to DNA shows good agreement to the proposed tetrameric model. B) The identified proximal, distal and submerged protoperithecia-1 (Sub-1) pseudo-palindromic DNA binding sequences in the *frq* promoter region (red text).

2.5 References

1. Ballario, P., et al. White collar-1, a central regulator of blue light responses in Neurospora, is a zinc finger protein. *EMBO J*. (1996). 15(7): 1650-7.
2. Linden, H. and G. Macino White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in Neurospora crassa. *EMBO J*. (1997). 16(1): 98-109.
3. Ballario, P. and G. Macino White collar proteins: PASsing the light signal in Neurospora crassa. *Trends Microbiol*. (1997). 5(11): 458-62.
4. Moglich, A., R.A. Ayers, and K. Moffat Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure*. (2009). 17(10): 1282-94.
5. Talora, C., et al. Role of a white collar-1-white collar-2 complex in blue-light signal transduction. *EMBO J*. (1999). 18(18): 4961-8.
6. Cheng, P., et al. WHITE COLLAR-1, a multifunctional neurospora protein involved in the circadian feedback loops, light sensing, and transcription repression of wc-2. *J Biol Chem*. (2003). 278(6): 3801-8.
7. Froehlich, A.C., J.J. Loros, and J.C. Dunlap Rhythmic binding of a WHITE COLLAR-containing complex to the frequency promoter is inhibited by FREQUENCY. *Proc Natl Acad Sci U S A*. (2003). 100(10): 5914-9.
8. Froehlich, A.C., et al. White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science*. (2002). 297(5582): 815-9.

9. Chen, C.H., et al. Genome-wide analysis of light-inducible responses reveals hierarchical light signalling in Neurospora. *EMBO J*. (2009). 28(8): 1029-42.

10. Ballario, P., et al. Roles in dimerization and blue light photoresponse of the PAS and LOV domains of Neurospora crassa white collar proteins. *Mol. Microbiol.* (1998). 29(3): 719-29.

11. Nambu, J.R., et al. The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell*. (1991). 67(6): 1157-67.

12. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. (1976). 72: 248-54.

13. Nielsen, S.S., M. Moller, and R.E. Gillilan High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell. *J. Appl. Crystallogr.* (2012). 45(Pt 2): 213-223.

14. Nielsen, S.S., et al. BioXTAS RAW, a software program for highthroughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J. Appl. Cryst.* (2009). 42: 959-964.

15. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* (1992). 25: 495-503.

16. Mylonas, E. and D.I. Svergun Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Cryst.* (2007). 40: 245-249.

17. Fischer, H., et al. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Cryst.* (2010). 43: 101-109.

18. Hinerman, J.M., J.D. Dignam, and T.C. Mueser Models for the binary complex of bacteriophage T4 gp59 helicase loading protein: gp32 single-stranded DNA-BINDING protein and ternary complex with pseudo-Y junction DNA. *J. Biol. Chem.* (2012). 287(22): 18608-18617.

19. Franke, D. and D.I. Svergun DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* (2009). 42: 342-346.

20. Volkov, V.V. and D.I. Svergun Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* (2003). 36: 860-864.

21. Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys Rev*. (2010). 2(1): 21-27.

22. Pettersen, E.F., et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. (2004). 25(13): 1605-12.

23. Svergun, D., C. Barberato, and M.H.J. Koch CRYSOL-a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *j Appl Crystallogr*. (1995). 28: 768-773.

24. Jacques, D.A., et al. Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta. Crystallogr. D Biol. Crystallogr.* (2012). 68(Pt 6): 620-626.

25. Hura, G.L., et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods*. (2009). 6(8): 606-612.

26. Jacques, D.A. and J. Trewhella Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. *Protein Sci.* (2010). 19(4): 642-657.

27. Putnam, C.D., et al. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* (2007). 40(3): 191-285.

28. Flaugh, S.L. and K.J. Lumb Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). *Biomacromolecules*. (2001). 2(2): 538-40.

29. Konarev, P.V., et al. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* (2003). 36: 1277-1282.

30. Petoukhov, M.V., et al. ATSAS 2.1 – towards automated and websupported small-angle scattering data analysis. *J. Appl. Cryst.* (2007). 40: 223-228.

31. Receveur-Bréchot, V. and D. Durand How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr. Protein Pept. Sc.* (2012). 13(1): 55-75.

32. Svergun, D.I. and M.B. Kozin Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* (2001). 34: 33-41.

33. King, H.A., et al. Structure of an enclosed dimer formed by the Drosophila period protein. *J. Mol. Biol.* (2011). 413(3): 561-72.

34. McRee, D.E. XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol*. (1999). 125(2-3): 156-65.

CHAPTER 3

FUTURE DIRECTIONS

3.1 Introduction

This section considers how additional efforts might be directed towards furthering our understanding of fungal circadian clock proteins, through development of a functional DNA binding assay, modification of purification conditions and design of new variants for biophysical characterization.

3.1.1 *N. crassa* White Collar I and II Zinc finger DNA Binding Studies

Preliminary experiments to demonstrate binding of White Collar (WC) I (705-1000), II (179-500) and copurified WC I/II (705-1000)/(179-500) zinc finger containing variants to rationally designed oligomers using fluoresce anisotropy and electrophoretic mobility shift assays with ethidium bromide visualized 2% agarose gels, were suggestive that binding was not occurring under conditions tested, as discussed in Appendix A. To address this issue, an alternative approach towards the rational design of the DNA oligomers might be useful.

The existing oligomers were designed with the thought that placing two conserved GATN motifs in close proximity might not only allow for DNA binding, but also help to order the proteins, thus facilitating their crystallization. A

redesign of the oligomers based variation in the number of included pseudo-palindromic sequence (red) for the identified proximal, distal and Sub-1 promoter sequences would allow for a variety of binding events (Figure 3.1). Additionally, 4 random bases on the 5' side of the identified sequence could also be included, which could act as a linker for subsequent fluorescent labeling experiments.

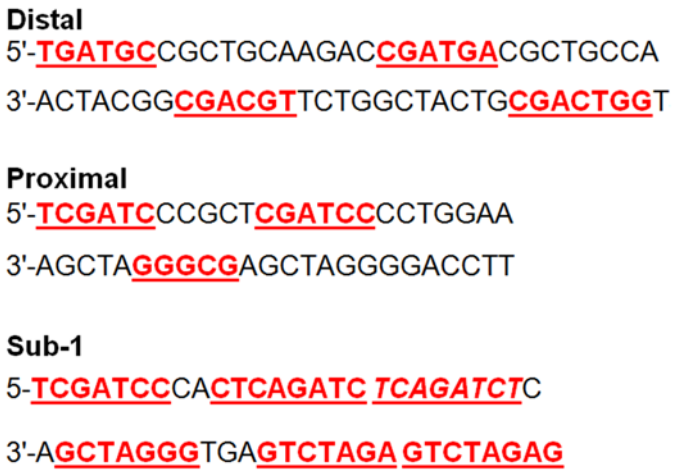

Figure 3.1 New Oligomers for DNA Binding Studies. DNA sequences incorporating the identified proximal, distal and Sub-1 binding sites are presented (red).

3.1.2 Experimental Variations for *N. crassa* White Collar II

Given the extensive crystallographic screening of *N. crassa* White Collar II proteins, coupled with the failure to obtain a crystal hit, additional perturbations to the existing approach are needed.

First, while significant progress was made in identifying stable WC II variants, relatively little success was made in the removal of the N-terminal His-tags, which might act to sterically hinder stabilizing interactions needed for

crystallization. Given the difficulty associated with tag cleavage as evidenced by SDS-PAGE gel analysis, expression with another affinity-tag such as small ubiquitin modifying protein (SUMO), which is reported to be reliably cleaved [1], might offer a reliable solution.

Another consideration is altering the buffer pH during the SEC purification step. All of the preparations described in Chapter 1 used a pH close to 7.5. The theoretical isoelectric point (pI) as calculated by the Scripps protein calculator: http://www.scripps.edu/cgi-bin/cdputnam/protcalc3 for WC II (179-500) is approximately 7.2 and 6.3 for each of the extended PAS only variants (174-455, 174-449, 174-453 and 174-455). Given this, the proteins are rendered with a slight negative net charge. As observed with *N. crassa* WC I/II (705-918)/(174-455) in Appendix A, significant variation in the amount of visible precipitation was observed in a buffer screen for proteins stored at room temperature. Preparation of individual WC variants could be performed with increased pH, which might serve to further stabilize the proteins in solution. Evaluation of the relative amount of aggregation from these screens could be monitored first visually as previously described, the followed by analytical SEC or with dynamic light scattering [2].

Reduction in sample aggregation would be particularly important for WC II (179-500), as results from small-angle X-ray scattering data (Chapter 2) indicated the protein was undergoing significant aggregation, thus preventing analysis. If

aggregation was reduced for WC II (179-500), the program Oligomer [3] could be used to generate form factors, which would allow for a calculation of the ratio of the WC I and II proteins in a the WC I/II complex at varying concentrations.

3.1.3 Experimental Variations for *C. neoformans* White Collar II

While SDS-PAGE and SEC analysis of *C. neoformans* WC II PAS variants discussed in Chapter 1 indicated excellent stability in solution, no crystal hits were obtained. To further address this issue, additional N- and C-terminal truncations designed to reduce variant length to the minimal PAS A core should be generated. As an alternative approach to methods previously taken for the identification of possible truncation points of the full-length protein, analysis of ordered/disordered regions were generated using the IUPred [4, 5] web server.

For IUPred analysis, an estimate of potential stabilizing inter-residue interactions for each amino acid in a 2-100 residues local environment were considered from the primary sequence. Once the presence or absence of these interactions was determined, assigned energies were totaled and averaged over a 21 residue window. A value of 0.5 is assigned as the cutoff between ordered and disorder in IUPred plots [4]. As seen in a long range disorder prediction for the *C. neoformans* WC II protein (Figure 3.2), the propensity for disorder is low from (41-155), which encompasses the Pfam predicted PAS domain (74-122). Immediately following this region, a significant spike in predicted disorder is

observed (155-172), with alternating values to approximately Glu262. Given the difficulty in estimating the C-terminal boundaries of the PAS domain, a potential screen of variants based on sharp changes in the raw data (not shown) suggest a screen starting at Leu41 and extending C-terminally to generate (41-138, 41-154, 41-172, 41-180, 41-214, 41-224 and 41-240).

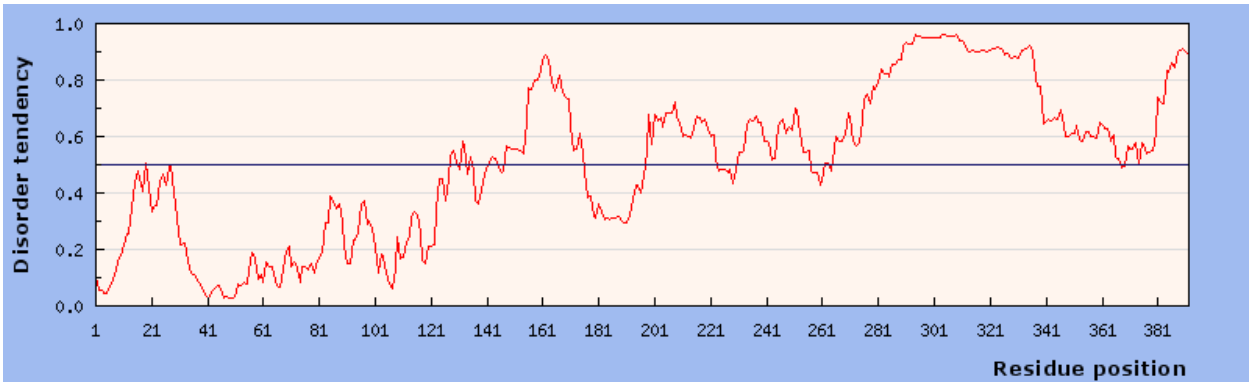

Figure 3.2 IUPRED Plot of Disorder *C. neoformans* White Collar II (1-392). An IUPred prediction of long range (global) disorder for the full-length *C. neoformans* WC II primary sequence shows a region of low disorder (41-154), followed by a spike for residues (155-172).

3.2 References

1. Arnau, J., et al. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif*. (2006). 48(1): 1-13.

2. Jancarik, J., et al. Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. *Acta Crystallogr D Biol Crystallogr*. (2004). 60(Pt 9): 1670-3.

3. Konarev, P.V., et al. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* (2003). 36: 1277-1282.

4. Dosztanyi, Z., et al. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. (2005). 21(16): 3433-4.

5. Dosztanyi, Z., et al. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol*. (2005). 347(4): 827-39.

APPENDIX A

SUPPLIMENTAL SECTION

Appendix A 1.1 Introduction

First addressed in the supplemental section are results from the purification and crystallographic screening of *N. crassa* Frequency (FRQ) and Frequency interacting RNA Helicase (FRH) variants. This is followed by tabular results from individual transcriptional activator purifications, initial results from DNA binding experiments and supplemental *N. crassa* WC I/II copurification experiments not addressed in Chapter 2.

Appendix A 1.1.1 Supplemental Results

Unless otherwise noted, expression and purification conditions for variants discussed in the supplemental section follow methods previously described in Chapter 1 for *N. crassa* White Collar II (179-500).

Appendix A 1.2 *N. crassa* Frequency

In an effort to further understand protein interactions in the *N. crassa* circadian clock, a short isoform of full-length FRQ (s-FRQ), which is important in clock temperature compensation [1], was generated for study. Additionally, a

variant based on a conserved N-terminal region containing a putative coiled coil domain comprised of amino acid residues (145-174) [2] was also tested.

Appendix A 1.2.1 Cloning *N. crassa* Frequency

S-FRQ, which starts at codon 100 in the *frq* open reading frame and continues to the C-terminal end as well as N-terminal constructs incorporating a putative coiled coil domain, (99-215) were cloned by coworkers using standard protocols into pet28 with an N-terminal 6xHis tag (Novagen), and provided for overexpression in *E. coli* BL21 (DE3) cells (Invitrogen).

Appendix A 1.2.2 *N. crassa* s-FRQ Purification and Crystallographic Screening

s-FRQ was grown at 37 °C followed by induction at 15 °C for 15 to 26 h. Following harvest and storage at -80 °C, protein was purified using IMAC followed by SEC at 4 °C (not shown) and concentrated to 1.3, 4.8 and 7.0 mg/mL.

As seen in SDS-PAGE analysis of concentrated s-FRQ, significant high and low molecular weight banding in addition to the 97 kDa theoretical molecular weight expected band is observed (Figure A 1.1).

In subsequent preparations, variation in induction temperature, addition of 1 mM phenylmethylsulfonyl fluoride (PMSF), cOmplete, EDTA-free protease

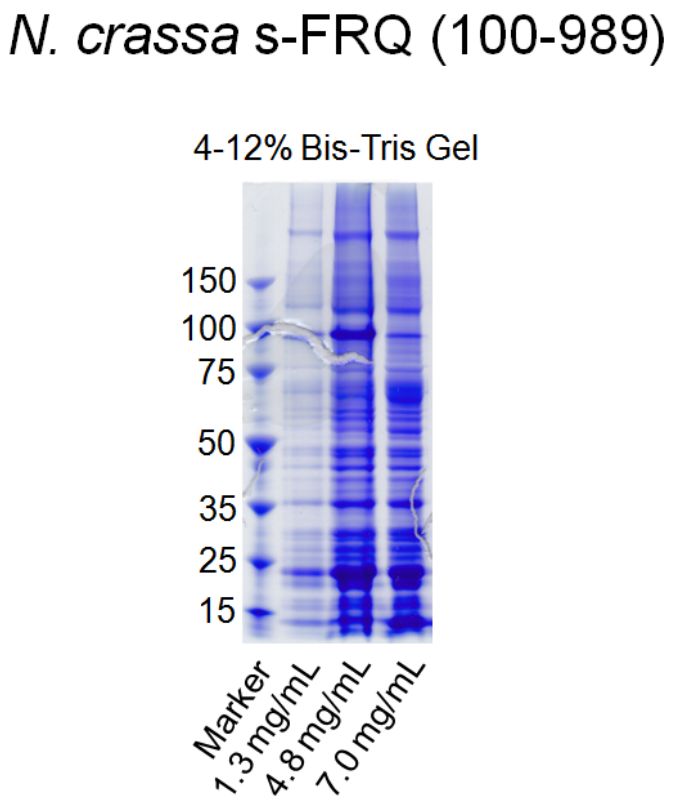

Figure A 1.1 Purification of a Short *N. crassa* Frequency Isoform Shows Significant Fragmentation. SDS-PAGE analysis of *N. crassa* (100-989) (s-FRQ) concentrated to 1.3, 4.8 and 7.0 mg/mL following IMAC and SEC purification show significant high and low molecular weight banding from preparations of s-FRQ isoform.

inhibitor cocktail tablets (Roche) and addition of 2 mM dithiothreitol (DTT) did not affect stability (not shown).

N. crassa full-length FRQ was manually screened using Crystal Screen 1 and 2 (Hampton Research) at 9 and 7.0 mg/mL respectively. Protein was also screened using Wizard Classic 1 and 2 (Emerald BioSystems) at 5.5 mg/mL. In all cases, the total drop volume was 2 µL with a 1:1 ratio of protein to well solution at 4 °C.

Additional screening was performed using an Art Robbins Phoenix with protein concentrated to 1 mg/mL in 1 µL drops (1:1 ratio of protein: well solution) for protein stored at 4, 17 °C and RT. All crystal screens were checked daily for three weeks, then weekly for approximately two months.

Appendix A 1.2.3 Purification and Crystallographic Screening of *N. crassa* N-terminal Frequency Variants

Analysis of SEC elution profile following IMAC purification of *N. crassa* FRQ (99-215) shows a principle elution peak eluting at 136 mL on a Superdex 75 prep grade gel filtration column (GE Healthcare) (Figure A 1.1).

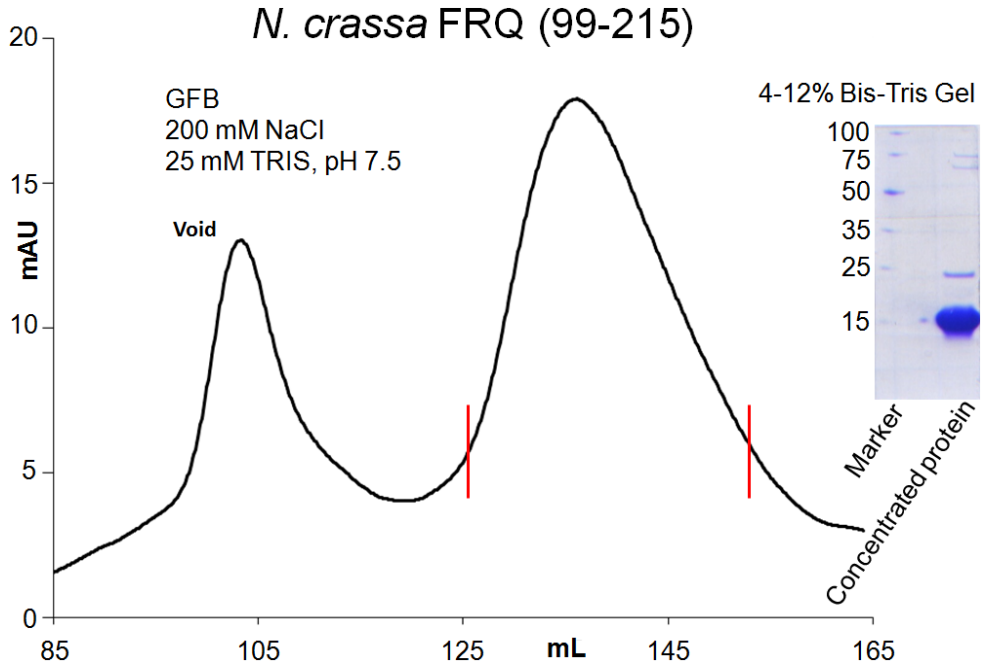

Figure A 1.1.1 Purification of a Conserved N-terminal Region of *N. crassa* Frequency. SEC elution profile for N. crassa FRQ (99-216) shows a primary elution peak at 136 mL. SDS-PAGE analysis of protein concentrated to 15 mg/mL following IMAC and SEC purification shows significant enrichment with no formation of high molecular weight bands or low molecular weight fragmentation.

Selected elution fractions, as denoted by red bars, were concentrated to 12 mg/mL and evaluated using SDS-PAGE gel analysis, which showed a 15 kDa band, consistent with the proteins 15 kDa theoretical molecular weight.

N. crassa FRQ (99-215) was manually screened at 20, 10, 15, 7.5 and 5 mg/L using Crystal Screen 1 and 2, Wizard Classic 1 and 2, PEG/Ion Screen 1 and 2 using a ratio of 1 µL protein to 1 µL well condition in hanging drops at RT. Crystal screens were checked daily for 1 month, then weekly for 3 months. The majority of well conditions (approximately 80%) from these screens was clear or had very light precipitation.

Appendix A 1.3 Cloning *N. crassa* Frequency

N. crassa FRH (114-1026) was cloned by coworkers using standard protocols into pet28 with an N-terminal 6xHis tag (Novagen), and provided for overexpression in *E. coli* BL21 (DE3) cells (Invitrogen).

Appendix A 1.3.1 Purification and Crystallographic Screening of *N. crassa* FRH

A 25 gram preparation of N. crassa FRH (114-1026) based on sequence homology to the structurally determined *Saccharomyces cerevisiae* RNA helicase Mtr4 (PDB: 3L9O) [3, 4] was grown at 37 °C for 3 h, followed by induction at 14 °C for approximately 20 h. Harvested protein was purified first using IMAC followed by SEC. As seen in SDS-PAGE analysis of protein

concentrated to 13 mg/mL (Figure A 1.2), a band consistent with the proteins 113 kDa molecular weight is observed. Also noted is minimal formation of high and low molecular weight bands indicative of protein aggregation and fragmentation respectively.

Concentrated protein was manually screened at 13 mg/L using Crystal Screen 1 and 2 and Wizard Classic 1 and 2 in 2 µL drops (1 µL protein:1 µL well condition) in hanging drops at 17 C. Crystal screens were checked daily for 3 weeks. Results from the screen indicated an equal distribution of clear to light, medium and heavy precipitated drops.

Appendix A 1.4 Experimental Notes for WC II and Related Homologs Purified in Chapter 2

Supplemental details and experimental results from the purification WC II homolog variants discussed in Chapter 1 are presented in Table A 1. Briefly, a qualitative estimate of the maximum total soluble protein (TSP) obtained for each variant, as determined by the Bradford assay of concentrated protein from IMAC capture followed by SEC purification is listed. Reported TSP values do not reflect target protein band in SDS-PAGE gel analysis and thus include impurities, aggregates and fragmentation products. Supplemental notes are provided for variants that readily precipitated during purification steps and appeared to have with the apparent molecular weight expected for a monomer).

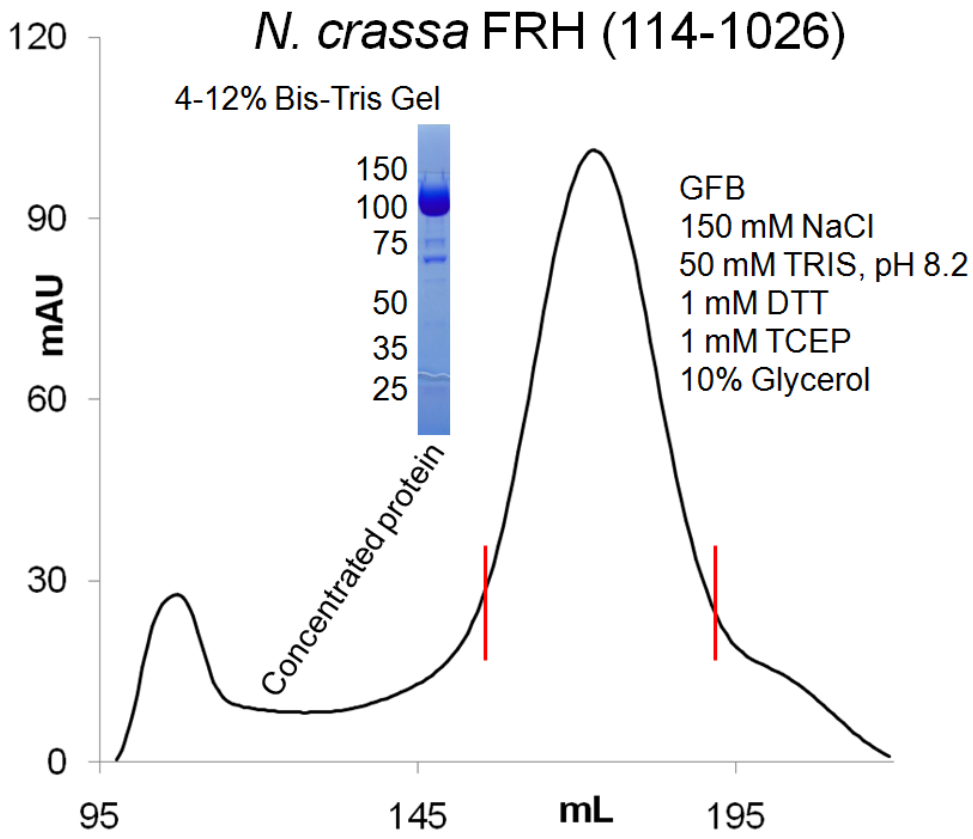

Figure A 1.2 SDS-PAGE Analysis of Purified *Neurospora crassa* Frequency Interacting RNA Helicase (FRH). SEC elution profile for *N. crassa* FRH (114-1026) shows a primary elution peak at 136 mL. SDS-PAGE analysis of protein concentrated to 13 mg/mL following IMAC and SEC purification shows significant enrichment with no observed formation of high molecular weight bands and minimal protein fragmentation.

successful cleavage of the His-tag as determined by target band shifts in SDS-PAGE analysis.

Also noted is the mM concentration of NaCl and Tris buffer in gel filtration buffer conditions (when varied). Finally a qualitative assessment of the SEC elution profile is provided with loss of an apparent dimer noted as either multimeric (multiple elution peaks) or monomeric (single elution value consistent

Table A 1 Summary of Experimental Notes for *N. crassa* WC II and Related Homologs Discussed in Chapter 1.

Variant	Max. Conc. (mg/mL)	Final Vol. (mL)	Total Protein (mg) /25 g Preparation	GFB (mM NaCl/ Tris buffer)	Elution Profile	Supplimental Notes
N. crassa						
(1-530)	-	-	-		-	No Soluble Protein
(154-285)	5.1	0.47	2.4	50/10	Multimeric	
(154-312)	0.7	0.50	0.4		Multimeric	Readily Precipitates
(154-323)	2	0.40	0.8		Multimeric	Readily Precipitates
(174-445)	12.9	0.60	7.7	50/10	Dimeric	
(174-449)	7	0.66	4.6	50/10	Dimeric	
(174-453)	12	0.44	5.3		Dimeric	
(174-455)	9.3	0.44	4.1	50/10	Dimeric	
(179-500)	14.2	0.74	10.5	50/10	Dimeric	His-tag Cleaved
P. blakesleeanus						
(1-354)	-	-	-		-	No Soluble Protein
(27-181)	7.6	1.00	7.6		Monomeric	Readily Precipitates
(27-286)	6.9	0.42	2.9		Dimeric	Readily Precipitates
(27-289)	5.2	0.47	2.4		Dimeric	Readily Precipitates
(27-293)	0.8	0.65	0.5		Multimeric	Readily Precipitates
(27-297)	2.1	0.65	1.4		Multimeric	Readily Precipitates
(27-300)	-	-	-		-	No Soluble Protein
(27-346)	-	-	-		-	No Soluble Protein
T. atroviride						
(1-484)	-	-	-		-	No Soluble Protein
(133-297)	-	-	-		-	No Soluble Protein
(133-318)	5	0.45	2.3		Multimeric	
(133-411)	8.3	0.50	4.2	50/10	Dimeric	
(133-415)	5.8	0.59	3.4		Dimeric	
(133-417)	9	0.61	5.5	50/10	Dimeric	
(133-467)	7	0.45	3.2	50/10	Dimeric	His-tag Cleaved
C. neoformans						
(1-392)	40	0.58	23.0	30/10	Dimeric	
(26-250)	67	0.56	37.5	30/10	Monomeric	His-tag Cleaved
(26-276)	70	0.17	11.9	50/10	Dimeric	His-tag Cleaved
(26-282)	24	0.50	12.0		Dimeric	
(26-289)	6.8	0.55	3.7		Dimeric	
(26-301)	22	0.75	16.5		Dimeric	
(26-306)	14.4	0.75	10.8		Dimeric	
(26-311)	12	0.60	7.2		Dimeric	
(26-315)	100	0.56	56.0		Dimeric	
(26-317)	30	0.57	17.1		Dimeric	
(26-320)	49	0.58	28.4		Dimeric	
(26-323)	34	0.75	25.5		Dimeric	
(26-329)	68	0.75	51.0		Dimeric	
(26-340)	30	0.58	17.4		Dimeric	
(26-378)	48	0.50	24.0		Dimeric	His-tag Cleaved
(26-380)	17	0.55	9.4		Dimeric	His-tag Cleaved
(26-383)	20	0.50	10.0		Dimeric	His-tag Cleaved

Appendix A 1.5 Results from Preliminary N. crassa White Collar II (179-500) DNA Binding Experiments

Efforts to develop a functional DNA binding assay for recombinantly expressed White Collar I and II transcriptional activators, containing the proteins C-terminal Zinc-finger, were based on a conserved imperfect GATA repeat (GATN) essential for induction of FRQ gene product as identified in deletion studies of proximal and distal regions in the *frq* promoter [5]. Also observed in this study, was identification of a three base-pair linker between the conserved GATN repeats from alignment of promoter region to another *N. crassa* light-induced clock gene product (albino-3).

From this work, rationally designed single-stranded DNA oligomers and their complimentary antisense sequences were generated and annealed using standard protocols, for use in both DNA binding assays and crystallographic screening. For crystallographic screening, design of the DNA oligomers focused on inclusion of the imperfect GATN repeat, where N can be any base, but must be the same in the repeat [5] (Figure 1.3). Further, observation that *N. crassa* WC II (179-500) appeared to elute as a dimer in SEC purifications drove the idea that two GATN repeats, either with (to potentially accommodate for steric hindrance during binding) or without separation by the identified albino-3 linker, would allow for simultaneous binding of both Zinc-fingers in the putative White Collar dimer, which might help in inducing order in crystallographic screening

experiments. Details for crystallographic screening with the DNA oligomer were detailed in Chapter 1.

Figure A 1.3 Rationally Designed DNA Oligomer for *N. crassa* White Collar Crystallographic and DNA Binding Experiments. Generation of a rationally designed single-stranded DNA oligomer for crystallographic screening included an imperfect GATA repeat (GATN), where N was selected as cytosine (black), separated with a three base-pair linker identified in the albino-3 coding region (green). An additional four base-pair linker (orange) was added to the on the 3' and 5' end of the sequence to facilitate linkage to a 5' Cy5 fluorescent label for use in fluorescent anisotropy experiments to assess White Collar DNA binding. Prior to use in crystallographic experiments the sense strand (top sequence) was annealed to the complimentary anti-sense strand (bottom sequence).

As a first approach in probing this interaction, a Cy5 fluorescent dye was added to the 5' end of the sense strand for use in fluorescence anisotropy experiments. The choice of Cy5 as a fluorescent reporter was determined by limitations in the range of measureable excitation/emission wavelengths in the experimental setup. Also considered was the possible steric hindrance of Cy5 on Zinc-finger/GATN binding. To address this issue a four-base pair linker was added in the sequence (Figure 1.3).

Subsequent fluorescent anisotropy experiments based on the µL titration of 143 µM *N. crassa* WC II (179-500 to a 25 nM Cy5 solution resulting in a final concentration ranging from 75 nM to 2.5 µM, failed to show changes in

fluorescence anisotropy. This result suggests that a binding event did not occur in the tested conditions (not shown).

As a second approach to investigating possible DNA binding, electrophoretic mobility shift experiments with ethidium bromide visualized 2% agarose gels was conducted following standard protocols. Briefly, 5 µL of *N. crassa* WC II (179-500), WC I (705-1000) and copurified WC I/II (705-1000)/(179-500) each at 6 µM, were mixed in 1:5, 2:5 and 5:1 ratios (protein:DNA) with the oligomers lacking the Cy5 label at 25 mM and allowed to incubate at 4 °C for 20-30 min. The mixtures were then loaded into a prechilled 2% agarose gel and run at 100 V for 30 min. Once complete, ethidium bromide was added and the gels were incubated with rocking for 30 min at room temperature (no EDTA was present in the gel or buffer). Results from these experiments clearly showed migration of the DNA in the gel with no shift observed for samples containing the protein relative to standards lacking the protein, suggestive that binding did not occur in tested conditions (not shown).

Appendix A 1.6 Supplemental *N. crassa* WC I/II Copurification Experiments

Based on observations from the copurification of *N. crassa* White Collar I/II proteins as described in Chapter 2, additional variants were generated for additional studies. Results presented reflect single preparations using conditions identified for N. crassa WC I/II (179-500)/(705-1000).

Appendix A 1.6.1 SEC-MALS Analysis of *N. crassa* WC I/II (174-455)/(705-918) in a High Ionic Strength/pH Buffer

As described in Chapter 2, low protein solubility at room temperature resulted in precipitation prior to SEC-MALS injection and thus, low signal intensity in the scattering experiment. To address this, copurified protein was screened against various buffers using Salt RX 1 and 2 (Hampton Research) at 6.5 mg/mL (GFB: 150 mM NaCl, 50 mM Tris buffer, pH 7.5, 5 mM DTT and 5 mM TCEP) in 2 µL total drop volumes (1 µL protein:1 µL well solution). Protein was stored at room temperature and monitored for visible precipitation after 8, 12, 24, 36 and 48 h. From this screen, NaCl was identified as having minimal visible precipitation upon inspection with a light microscope. Using the same method, the protein was rescreened against NaCl concentrations ranging from 50 to 950 mM in 50 mM increments. Results from this screen suggested reduced visible precipitation in concentrations at and above 300 mM NaCl. As a final variation, the protein was again rescreened in GFB containing 300 mM NaCl, 50 mM Tris buffer, pH 7.5, 5 mM DTT and 5 mM TCEP against a pH screen from 2 to 11 in 0.5 increments using a variety of buffers. From these results, drops at and above pH 9.0 did not display visible precipitation at room temperature even after 3 days.

With these results in mind, a 25 gram preparation of *N. crassa* WC I/II (174-455)/(705-918) was subjected to IMAC and SEC purification. As seen in Figure A 1.3.1, the SEC elution profile using a GFB containing 300 mM NaCl, 50

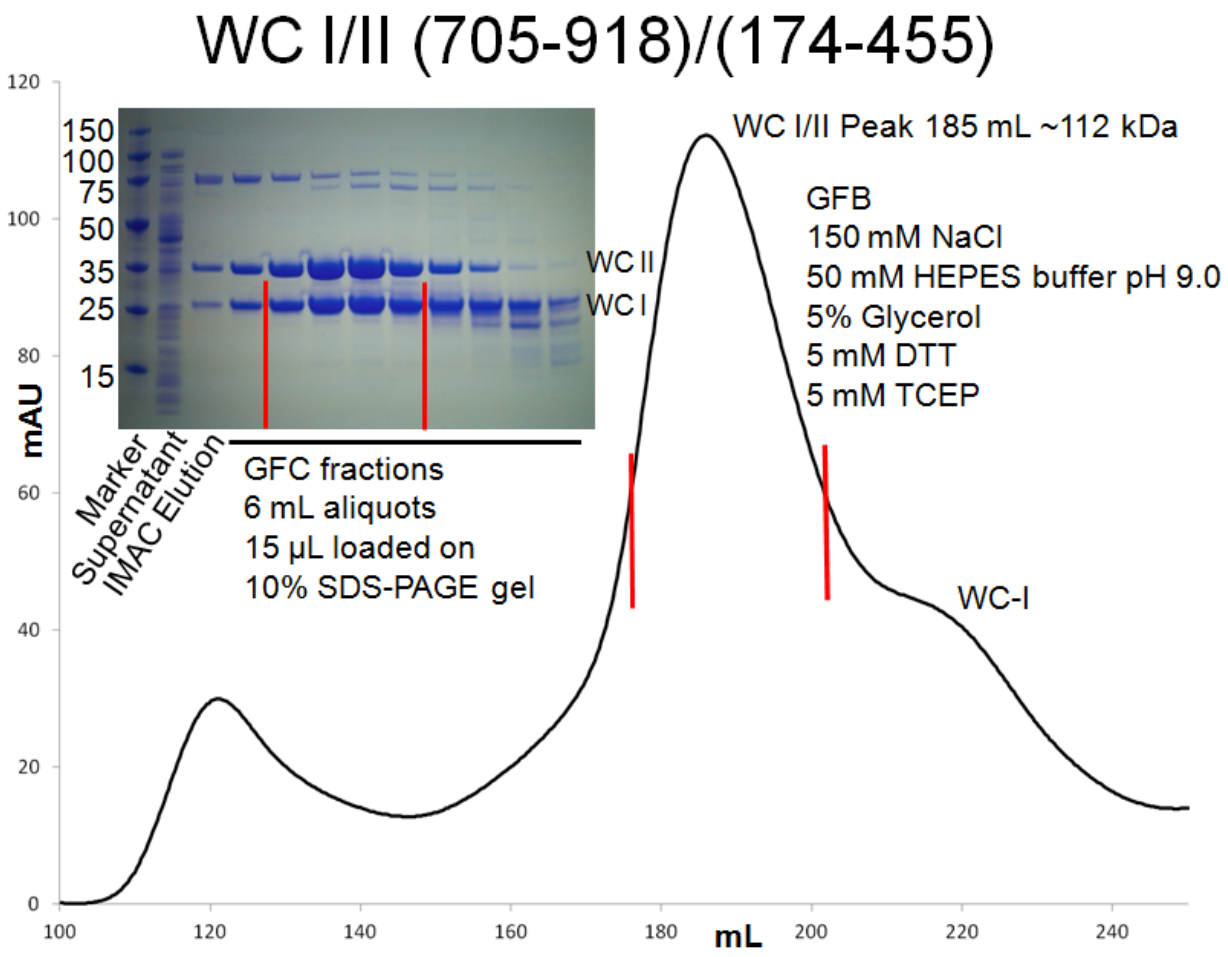

Figure A 1.3.1 ***N. crassa*** **WC I/II (174-455)/(705-918) Shows a Decreased Hydrodynamic Radius at pH 9.0.** The SEC elution profile for *N. crassa* WC I/II (174-455)/(705-918) shows a primary elution peak at 185 mL, suggestive of a 112 kDa species. SDS-PAGE analysis of SEC elution fractions show bands corresponding to the WCI and II theoretical molecular weight, with minimal protein fragmentation.

mM HEPES, pH 9.0, 5 mM DTT and 5 mM TCEP showed a principle elution peak at 185 mL, which when compared to known standards suggested a species with a 112 kDa molecular weight, small than the proteins 118 kDa theoretical molecular weight. This value is also smaller than the 130 kDa observed previously using a 150 mM NaCl and Tris buffer pH 7.5 buffer, suggestive of a

small hydrodynamic radius for the complex. SDS-PAGE gel analysis of SEC elution fractions indicated the presence of high molecular weight bands, which are suggestive of irreversible protein aggregation. From analysis of the preparation it is unclear if the increased high molecular weight banding is due to protein handling or reduced stability in the buffer.

Following SEC purification, selected fractions, as denoted by red bars on the SEC elution profile and accompanying SDS-PAGE gel were concentrated to 7.5 mg/mL and stored at 4 °C for SEC-MALS analysis using the same experimental procedure described in Chapter 2 methods section.

Prior to injection, the concentrated protein was stored at room temp and did no visibly precipitate prior to injection. Following injection, SEC-MALS analysis (Figure A 1.3.2) indicated a species with a 103 kDa (±2%) molecular weight, smaller than the theoretical molecular weight for a WC I/II complex, as well as the 118 kDa value from the previous SEC-MALS analysis (Figure 2.3.3.2).

Further observed is a decrease in molecular weight for the complex during protein elution, possibly suggestive of dissociation of the complex during the experiment, which was not observed in the previous SEC-MALS experiment.

Finally, noted was the observation that the protein appeared to interact with the column during the experiment, as evidenced by a low scattering and

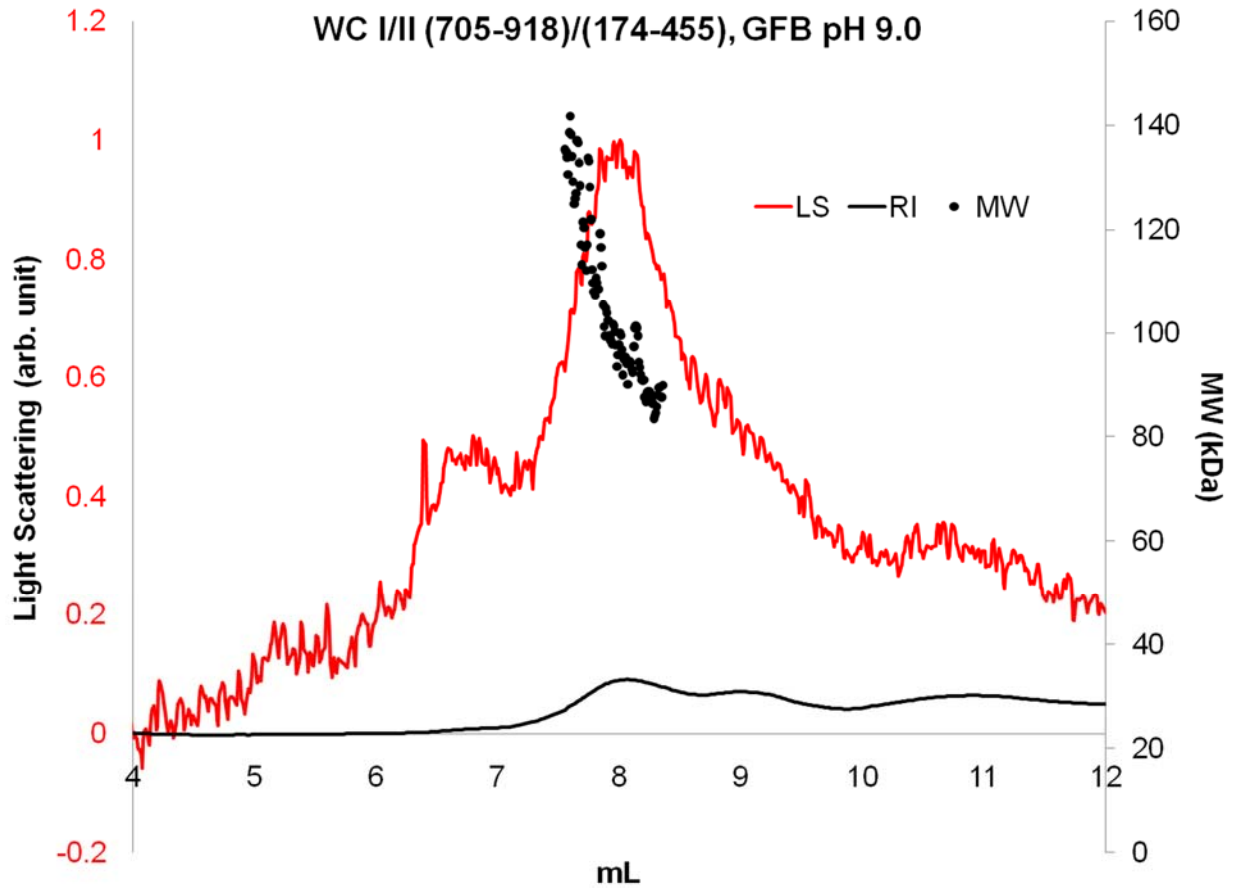

Figure A 1.3.2 *N. crassa* WC I/II (174-455)/(705-918) SEC-MALS Analysis. SEC-MALS analysis of copurified of *N. crassa* WC I/II (174-455)/(705-918) in 300 mM NaCl and HEPES pH 9.0, resulted in a 103 (±2%) kDa complex. This value is smaller than the complexes theoretical 118 kDa molecular weight.

refractive index signal persistent (not shown).

Appendix A 1.6.2 Copurification of *N. crassa* WC I (705-918)/VVD

As described in Chapter 2, an 37 amino acid N-terminally truncated variant of *N. crassa* Vivid (VVD) was copurified with *N. crassa* WC I (705-918) using conditions previously described in Chapter 2 for *N. crassa* WC I/II (174-455)/(705-918).

SDS-PAGE analysis of SEC elution fractions from a 25 g preparation (Figure A 1.3.3) failed to show coelution of the two proteins principle elution peaks, suggesting a lower affinity between the proteins in the current buffer composition.

SEC elutions values for the WC I (705-918) and VVD were consistent with those previously observed using a HiLoad 26/60 Superdex 75 prep grade gel filtration column (GE Healthcare) equilibrated with 150 mM NaCl, 50 mM Tris pH 7.5, 10% v/v glycerol, 5 mM DTT and 5 mM TCEP (not shown).

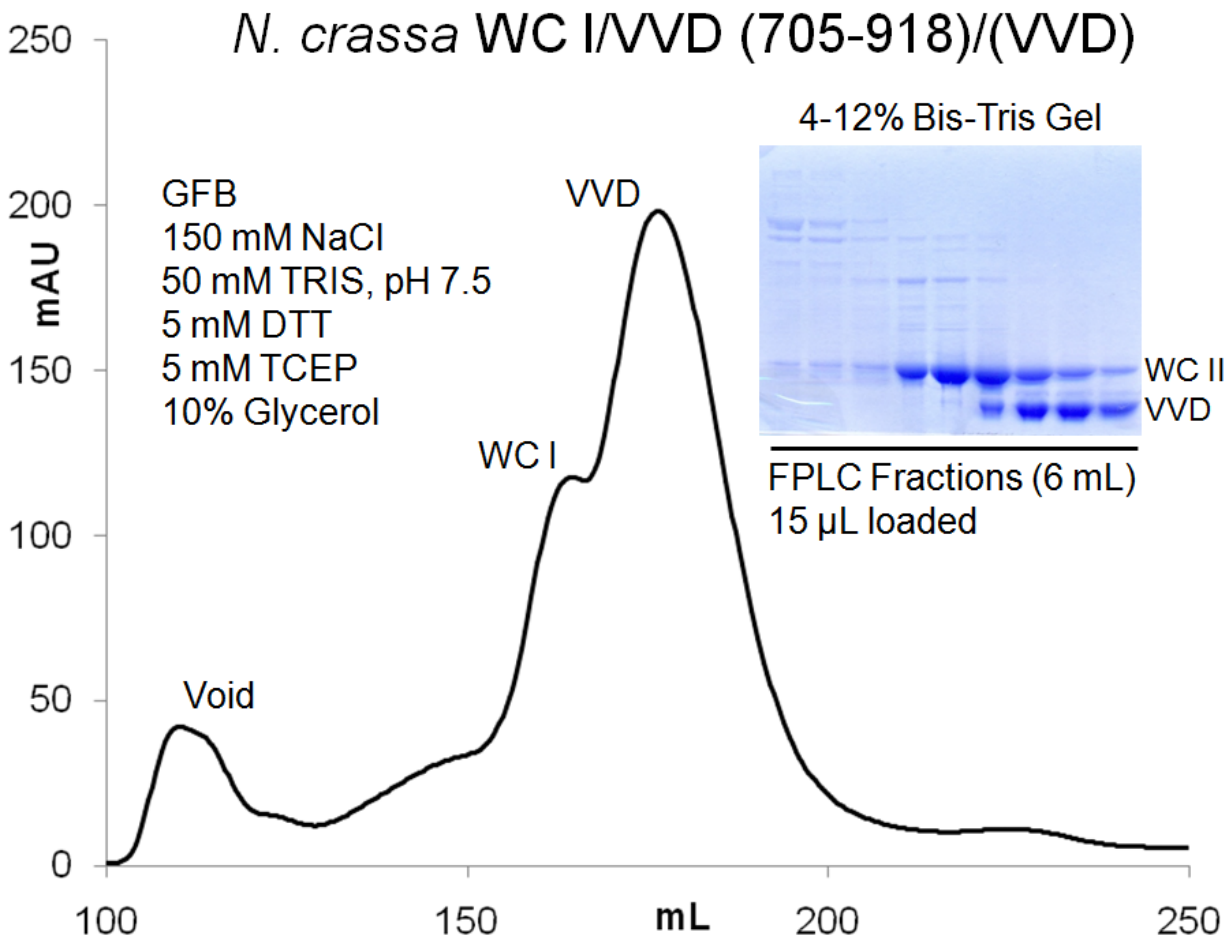

Figure A 1.3.3 Copurification of *N. crassa* WC I (705-918) and Vivid Did Not Show Coelution in SEC Purification. SDS-PAGE gel analysis and SEC elution fractions of co-purified *N. crassa* WC I (705-918)/VVD, does not shown co-elution in SEC purification, indicating a weak nor non-existent interaction for the two proteins in tested buffer conditions.

Appendix A 1.7 References

1. Akman, O.E., et al. Isoform switching facilitates period control in the Neurospora crassa circadian clock. *Mol Syst Biol*. (2008). 4: 164.

2. Cheng, P., et al. Coiled-coil domain-mediated FRQ-FRQ interaction is essential for its circadian clock function in Neurospora. *EMBO J*. (2001). 20(1-2): 101-8.

3. Jackson, R.N., et al. The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J*. (2010). 29(13): 2205-16.

4. Weir, J.R., et al. Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc Natl Acad Sci U S A*. (2010). 107(27): 12139-44.

5. Froehlich, A.C., et al. White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science*. (2002). 297(5582): 815-9.